Differential RNA-Seq and transcription start site annotation in *Chlamydia*

Dissertation

To Fulfill the
Requirements for the Degree of

doctor rerum naturalium (Dr. rer. nat.)

Submitted to the Council of the Faculty
of Biological Sciences
of the Friedrich Schiller University Jena

by
Thomas Beder (M.Sc.)
born on July 17, 1988 in Weimar

Jena, 2019

Table of Contents

Abbreviations

	5’P	5’-monophosphate		5’PPP	5’-triphosphate
	°C	Degree Celsius		Δ	Delta
	µg	Microgram		µl	Microliter
	µM	Micromolar		%	Percent
A	asRNA	Antisense RNA	**B**	bp	Base pair
C	*C.*	*Chlamydia*		Ca	Calcium
	cDNA	Complementary DNA		CDS	Coding DNA
	Cl	Chlorine		CO2	Carbon dioxide
	Ct	Cycle threshold	**D**	ddH2O	Double-distilled water
	DEG	Differential expressed gene		DMEM	Dulbeccos’s modified Eagle medium
	DNA	Desoxyribose nucleic acid		dRNA-Seq	Differential RNA sequencing
E	EB	Elementary body		*E.*	*Escherichia*
	EDTA	Ethylenediaminetetraacetic		e.g.	For example
		Acid	**F**	FBS	Fetal bovine serum
G	gDNA	genomic DNA	**H**	Hcl	Hydrochloric acid
	hpi	Hours post infection	**I**	i.e.	That is
	IFN-c	Interferon-gamma		IFU/ml	Inclusion forming units per ml
	Inc	Inclusion membrane protein		K2HPO4	Dipotassium hydrogen phosphate
L	l	Liter		LGV	Lymphogranuloma venereum
M	Mbp	Megabase pairs		ml	Mililiter
	Mg	Magnesium		MOI	Multiplicity of infection
N	NaCl	Sodium chloride		ncRNA	non-coding RNA
	NGS	Next-generation-sequencing		nt	Nucleotide
O	OMC	Outer membrane complex		Omp	Outer membrane protein
	ORF	Open reading frame	**P**	PBS	Phosphate buffered saline
	PCR	Polymerase chain reaction		Pmp	Polymorphic membrane protein
	PRC	Precision		PZ	Plasticity zone
Q	qPCR	Real-time quantitative PCR	**R**	RB	Reticulate body
	RCF	Relative centrifugal force		RNA-Seq	RNA sequencing
	rRNA	Ribosomal RNA		RT	Room temperature
	RT-qPCR	Reverse transcriptase qPCR	**S**	Sn	Sensitivity
	SP	Specificity		SPG	Sucrose-phosphate-plutamate transport medium
	sRNA	small regulatory RNA	**T**	T3SA	Type III secretion system apparatus
	T3SE	Type III secretion system effectors		TEM	Transmission electron microscope
	tRNA	transfer RNA		TSSs	Transcription start site
U	U	Enzyme unit		UTRs	Untranslated region
V	VF	Virulence factor		v/v	Volume/volume
W	w/v	Weight/volume			

1 Summary

Chlamydiae are obligate intracellular Gram-negative bacteria that infect a wide range of hosts. However, host specificity and virulence of the individual species differ drastically even though all members of the phylum share the majority of genes which are encoded on a much reduced genome. Striking examples of this contrast are seen in the close relatives *Chlamydia* (*C*.) *psittaci* and *C*. *abortus*. *C*. *psittaci* is the causative agent of psittacosis, the most widespread zoonotic chlamydiosis, while the less widespread *C*. *abortus* is still of clinical relevance because of its ability to colonize the human placenta. Other chlamydia-like organisms, such as *Waddlia* (*W*.) *chondrophila*, also occasionally infect humans, cause similar symptoms and exhibit the *Chlamydia*-specific biphasic developmental cycle.

In this work, differential RNA-Sequencing (dRNA-Seq) was applied to purified infectious and non-infectious developmental forms of *C*. *psittaci*, *C*. *abortus* and *W*. *chondrophila*. The comparison revealed that both, infectious and non-infectious states contain distinct transcriptomes that match their biological functions. Furthermore, considerable differences in the expression of homologous virulence factors were found. A second focus of this work was the reliable annotation of transcription start sites (TSSs) using dRNA-Seq data. The presented approach is based on optimization and combination of three TSS annotation tools and its success was validated with data from *Escherichia* (*E*.) *coli* for which besides the dRNA-Seq another high throughput data set (Cappable-seq) for TSS annotation was available. In total 9,802 TSSs with high precision were identified based on data of *E*. *coli*, *C. pneumoniae*, *C*. *psittaci*, *C*. *abortus* and *W*. *chondrophila*. Knowing the precise TSSs positions, the corresponding 5’ untranslated regions (UTRs) and conservation among species could be analyzed. This revealed that among *Chlamydia*, positions of TSSs and 5’UTR sequences are evolutionary less conserved than the corresponding coding sequences. This is fascinating because the coding sequences of *Chlamydia* are conserved and it has been indicated that the differing expression levels of transcripts, rather than sequence dissimilarities alone might contribute to the differences in host specificity and tissue tropism. In concordance, varying TSSs positions, 5’UTR and promoter sequences may contribute to the distinct expression levels observed for highly conserved genes. Furthermore, a promoter motif analysis was performed based on the TSS annotations and 2,157 chlamydial promoters (60.57%) with conserved motifs were found most of which contain potential σ factor binding sites. In summary, the work describes a gene expression analyses of EB and RBs and the reliable annotation of TSSs using dRNA-Seq data.

2 Zusammenfassung

Chlamydiae sind obligat intrazelluläre Gram-negative Bakterien die eine Vielzahl von Wirtsorganismen befallen. Alle Arten des Phylums besitzen ein reduziertes Genom und die Mehrzahl der proteinkodierenden Gene ist konserviert. Trotz dieser Gemeinsamkeiten sind Wirtsspezifität und Gewebetropismus der einzelnen Spezies stark verschieden. Exemplarische hierfür sind *Chlamydia* (*C*.) *psittaci* und *C*. *abortus*. *C*. *psittaci* ist der Erreger der Psittakose (Papageienkrankheit), welche die weitverbreitetste zoonotische Form der Chlamydiose darstellt. *C*. *abortus* Infektionen sind weniger häufig, aber dennoch von klinischer Relevanz wegen der Fähigkeit des Erregers die Plazenta zu befallen. Neben den „klassischen" Spezies die alle zur Familie der *Chlamydiaceae* gehören wurden in den letzten Jahrzehnten verschiedene neue Arten entdeckt. Einige von ihnen, wie z.B. *Waddlia* (*W*.) *chondrophila*, können auch den Menschen befallen und die Krankheitsbilder ähneln den der *Chlamydiaceae*. Eine weitere Gemeinsamkeit aller Chlamydien ist der charakteristische biphasische Entwicklungszyklus in dem infektiöse Elementarkörperchen (EK) und Retikularkörperchen (RK) wechseln.

In der vorliegenden Arbeit wurden die Transkriptome der EK und RK analysiert. Dazu wurde eine Methode namens differentielle RNA Sequenzierung (dRNA-Seq) verwendet und ein Vergleich von *C*. *psittaci*, *C*. *abortus* und *W*. *chondrophila* zeigte, dass beide Formen große Unterschiede in der Genexpression aufweisen. Des Weiteren sind viele homologe Virulenzfaktoren in den Spezies unterschiedlich stark exprimiert, was zu den verschieden Krankheitsbildern beitragen könnte. Ein zweiter Schwerpunkt der Arbeit lag auf der präzisen Bestimmung der Tanskriptionsstartseiten (TSSs) mithilfe der dRNA-Seq Daten. Zu diesem Zweck wurden die Vorhersagen drei verschiedener Programme optimiert und kombiniert. Diese Herangehensweise konnte mit Daten von *Escherichia* (*E*.) *coli* validiert werden, da neben den dRNA-Seq auch ein weiterer Datensatz (Cappable-seq) zum Abgleich zur Verfügung stand. Insgesamt konnten auf diese Weise 9.802 TSSs mit großer Genauigkeit in *E*. *coli*, *C. pneumoniae*, *C*. *psittaci*, *C*. *abortus* und *W. chondrophila* bestimmt werden. Des Weiteren, war es möglich durch die genaue Lokalisation der TSSs die 5' untranslatierten Bereiche (5'UTRs) und deren Konservierung zu analysieren. Dadurch wurde gezeigt, dass die TSS Positionen und 5'UTRs weniger stark konserviert sind als die dazugehörigen kodierenden Sequenzen. Diese Differenz könnte zu den Expressionsunterschieden stark konservierter Gene zwischen den Spezies beitragen. Ferner wurden in 2.157 Promotoren (60,57%) konservierte Motive gefunden von denen die meisten potenzielle σ-Faktor Bindestellen sind. Zusammengefasst beschreibt die vorliegende Arbeit die Analyse der EK und RK Transkriptome und eine präzise TSS Bestimmung.

3 Introduction

3.1 *Chlamydia* – protozoans, viruses or bacteria?

Chlamydiae is a unique phylum of bacteria both, historically and with regard to various biological aspects. The term "*Chlamydia*" is derived from a notion of Ludwig Halberstaedter and Stanislaus von Prowazek who first described the intracellular organisms as early as 1907 and believed to had discovered "mantled protozoans" (Halberstaedter and Prowazek, 1907). The two scientists investigated stained conjunctival scrapings from trachoma patients in which the intracellular vacuole, that contains numerous *Chlamydia*, appeared as a single organism. After their discovery, pathogens forming similar intracellular inclusions were found in samples obtained from patients with lymphogranuloma venereum (LGV), a sexually transmitted infectious disease, and during a worldwide pandemic of an atypical and acute pneumonia, which resulted from contact with psittacine birds (Durand, Nicolas and Favre, 1913; Bedson, Western and Simpson, 1930; Coles, 1930; Lillie, 1930). In 1935, the pathogens had been classified as viruses, as, because of their small size (infectious states ~0.3 µm), they could be passed through bacterial filters and because it was not possible to grow them on artificial media (Miyagawa *et al.*, 1935). Ascertained in different human tissues and associated with different conditions it was not until 1942 that Rake and Jones showed that trachoma and conjunctivitis organisms were much like those of the psittacosis-LGV group, by identifying a common complement-fixing antigen (Rake and Jones, 1942). The hypothesis that *Chlamydia* are viruses lasted until the establishment of electron microscopy when the organisms were classified as bacteria, essentially because they possess both DNA and RNA, ribosomes and have a cell wall similar to that of Gram-negative bacteria (Moulder, 1966). The difficulties in classification of *Chlamydia* because of their small size and intracellular life style and their association with various diseases in human and animals already indicate the uniqueness of these organisms.

3.2 *Chlamydia* cause a wide range of infectious diseases

Chlamydia are energy parasites found in a variety of animals ranging from protozoa, mollusks, insects, fish, reptiles, birds, marsupials up to mammals such as humans (reviewed in Horn, 2008; Borel, Polkinghorne and Pospischil, 2018). The species first described by Halberstaedter and Prowazek was *Chlamydia* (*C*.) *trachomatis*, the most prominent chlamydial representative that causes trachoma as well as sexually transmitted infections (**Table1**). Trachoma mainly occurs in tropical regions and is the world's leading infectious cause of irreversible blindness (Rohde *et al.*, 2010; Flaxman *et al.*, 2017). *C*. *trachomatis* is also the most common cause of sexually transmitted diseases, with estimated 128 million new cases every year (Newman *et al.*, 2015). Acute infections lead to conditions like cervicitis, urethritis, and genital ulceration. Repeated and chronic infections can result in long term sequelae including pelvic inflammatory disease, ectopic pregnancy up to infertility (Global Burden of Disease Study 2013 Collaborators, 2015). Moreover, an association between human miscarriage and *C*. *trachomatis* infections has been demonstrated based on serologic and molecular evidence (Baud *et al.*, 2011). In summary, the clinical relevance makes *C*. *trachomatis* the best studied chlamydial representative.

The second major human pathogen among the *Chlamydia* is *C*. *pneumoniae* a widespread respiratory pathogen (**Table1**) and a causative agent of community-acquired pneumonia (Kuo *et al.*, 1995). Chronic infections are associated with an enhanced risk of developing atherosclerosis (Kuo *et al.*, 1993), chronic lung disease (Hahn, Dodge and Golubjatnikov, 1991) and Alzheimer's disease (Balin *et al.*, 1998). This shows that *C*. *pneumoniae* can infect various tissues and disseminates within the human body. The significance of this pathogen in public health is also emphasized by the high prevalence since antibodies against *C*. *pneumoniae* are developed by most adults within their lifetime (Blasi *et al.*, 1998). Another interesting difference to *C*. *trachomatis* is that *C*. *pneumoniae* was found in various animal species, meaning that *C*. *pneumoniae* has a

wide host range, whereas *C*. *trachomatis* is a strictly human pathogen (reviewed in Horn, 2008).

3.3 Primary animal *Chlamydia* with zoonotic potential

Besides the major human pathogens *C*. *trachomatis* and *C*. *pneumoniae* there exist other chlamydial species predominantly infecting animals. Yet some of these species may be transmitted to humans (**Table 1**) in a process called zoonosis. The most prominent zoonotic *Chlamydia* are *C*. *psittaci* and *C*. *abortus* with well studied modes of transmission to humans and distinct clinical manifestation (Longbottom and Coulter, 2003). *C*. *psittaci* is the causative agent of psittacosis (also known as ornithosis or parrot-fever), a disease first recognized in Switzerland and associated with contact to tropical birds in 1879 (Ritter, 1880). Later, *C*. *psittaci* was indicated as the aetiological agent of the disease (Bedson, Western and Simpson, 1930). Ornithosis is the most widespread zoonotic chlamydiosis (Beeckman and Vanrompay, 2009; Knittler *et al.*, 2014; Lagae *et al.*, 2014) and symptoms in humans range from clinically silent or flu-like to an acute illness with severe pneumonia and death (Crosse, 1990; Petrovay and Balla, 2008). Even human-to-human transmissions of *C*. *psittaci* have been reported, where a patient became severely ill after exposure to wild bird droppings followed by transmission of the infection to ten other people, including hospital staff (Wallensten, Fredlund and Runehagen, 2014). *Ab initio* psittacosis was associated with birds and till today *C*. *psittaci* has been found in more than 100 different bird species (reviewed in Horn, 2008). During the last decades, psittacosis outbreaks were reported in poultry farms all over the world with great financial losses and high risks for workers since the zoonotic transfer occurs through inhalation of contaminated aerosols originating from feathers or feces (Lagae *et al.*, 2014). The symptoms of avian chlamydiosis include lethargy, hyperthermia, abnormal excretions and respiratory distress (Knittler *et al.*, 2014). In addition to infections in birds, *C*. *psittaci* was demonstrated as the causative agent of equine abortions and a risk of zoonotic transmission originating from infected

placental membranes has been indicated (reviewed in Borel, Polkinghorne and Pospischil, 2018).

The closest relative of *C*. *psittaci* is *C*. *abortus*, which is less dangerous but still of economic and medical importance. *C*. *abortus* is the aetiological agent of ovine enzootic abortion, the major cause of infectious lamb loss in many countries world-wide (Essig and Longbottom, 2015). Although the organism is principally endemic to sheep and goats (particularly in lowland flocks), infections of other host species, such as cattle, pigs, horses or deer can occur (Longbottom and Coulter, 2003) and even *C*. *abortus* infections in Chinese yaks have been reported (Chen *et al.*, 2014). However, the specific extend of *C*. *abortus* infections in live stocks is most likely still underrated due to difficulties in culturing and thereby diagnosis of the pathogen (Vidal *et al.*, 2017). In sheep flocks, the prevalence of *C*. *abortus* can be estimated to be very high. Lenzko *et al.*, tested more than 50% of the flocks in Thuringia (Germany) positive for *C*. *abortus* even though the sheep flock appeared clinically healthy (Lenzko *et al.*, 2011). *C*. *abortus* may be inhaled by people exposed to the animals, which becomes medically important because of the pathogen's ability to colonize the human placenta (Longbottom and Coulter, 2003). Therefore, the major threat concerns pregnant women with close contact to stillborn ruminants and resulting infections may lead to preterm stillbirth and sepsis-like symptoms (Walder *et al.*, 2005). Several other chlamydial species that infect animals are known (**Table 1**). Of these *C*. *felis*, *C*. *caviae* and *C*. *suis* may cause infrequent human infections with various clinical presentations but currently only *C*. *felis* is implicated with compelling evidence to be transmitted to humans and causing conjunctivitis (Hartley *et al.*, 2001).

3.4 Environmental *Chlamydia*

Besides the “classic” chlamydial species, all of which belong to the family *Chlamydiaceae*, there exist various *Chlamydia*-like pathogens including members of the families *Parachlamydiaceae*, *Simkaniaceae* and *Waddliaceae* (**Table1**). 1995, the first

Chlamydia-like organism has been described as a laboratory cell-culture contaminant and suggested as a potential novel genus of the family *Chlamydiaceae* based on its 16S-DNA sequence (Kahane, Metzer and Friedman, 1995). Later this species was classified as a member of a new family *Simkaniaceae*. Cell-culture experiments suggested that *Simkania negevensis* may be associated with respiratory diseases, but the organism has not been detected in humans so far (Vouga, Baud and Greub, 2017). Besides *Simkania negevensis* several other chlamydia-like organisms have been found in the past two decades, most of which are proposed to be parasites of amoebae (Amann *et al.*, 1997). This assumption is supported by the fact, that in cases where the natural host remained unknown these bacteria were successfully recovered upon co-cultivation with amoebae (Collingro *et al.*, 2005; Thomas, Casson and Greub, 2006). For some environmental *Chlamydia*-like species such as *W. chondrophila* there is evidence for pathogenicity (reviewed in Borel, Polkinghorne and Pospischil, 2018). This species has been associated with respiratory disease and miscarriage in humans and can be cultured in mammalian cells (Dilbeck *et al.*, 1990; Henning *et al.*, 2002; Dilbeck-Robertson *et al.*, 2003; Haider et al., 2008; Goy and Greub, 2009; Baud, D. Baud *et al.*, 2011, Greub and Centre, 2011; Barkallah et al., 2013; Baud, Osterheld, et al., 2014). Thus, the disease outcome of *W. chondrophila* closely resembles the diseases caused by *C. psittaci* and *C. abortus* (**Table 1**). This makes studying *W. chondrophila* as an environmental counterpart of the medically important *Chlamydiaceae* interesting, as it represents a possibility to further define characteristic features of chlamydial pathogenicity and host adaptation. Both known *W. chondrophila* strains were isolated from aborted bovine fetuses (Dilbeck *et al.*, 1990; Henning *et al.*, 2002) and thus it was early proposed that *W. chondrophila* might cause abortion in ruminants. In this context, serological testing of bovine sera showed higher *W. chondrophila* titer in cattle that had aborted (Dilbeck-Robertson *et al.*, 2003) and an association of *W. chondrophila* with bovine abortion was also demonstrated by the detection of the pathogen in veterinary samples from Tunisia (Barkallah *et al.*, 2013).

Table 1. Diseases caused by *Chlamyida* in man an in animals (adapted from Rohde *et al.*, 2010 and Borel *et al.*, 2018)

Genus	Species	Diseases in man	Frequency of infections in man	Diseases in animals
Chlamyida	*abortus*	Systemic infection Abortion	Rare	Enzootic ovine abortion, also Occasionally abortion in cattle
	caviae	Non-pathogenic		Conjunctivitis (guinea pig)
	felis	Conjunctivitis	Rare	Keratoconjunctivitis (cat)
	muridarum	Non-pathogenic		Pneumonitis (mice and hamster)
	pecorum	Non-pathogenic		Enteritis, abortion, conjuncti vitis, pneumonia, Encephalomyelitis, polyarthritis (ruminants)
	pneumoniae TWAR biovar	Respiratory infections Atypical pneumonia Reactive arthritis Infestation of the arteriosclerotic plaque	Very frequent Occasional Occasional Probably frequent	Unknown
	psittaci	Respiratory infections Pneumonia Myocarditis Hepatitis Encephalitis	Occasional Occasional Rare Very rare Very rare	Psittacosis/ornithosis (true parrots, domestic poultry)
	trachomatis	Urogenital infections Adnexitis Reactive arthritis Swimming pool conjunctivitis Neonatal conjunctivitis Neonatal pneumonia	Frequent in adolescents Occasional Occasional Rare Rare Rare	Unknown
	trachomatis, trachoma biovar	Trachoma	Frequent in tropical countires	Unknown
	trachomatis, LGV biovar	Venereal lymphogranuloma	Occasional in tropical countries Also in HIV- positive men in Europe	Unknown
	suis	Non-pathogenic		Pneumonia, enteritis, conjunctivitis (pig)
Simkania	*negevensis*	Unknown	Unknown	Unknown
Parachlamydia	*acantamoebae*	Unknown	Unknown	Abortion, pneumonia (bovines, sheep, goats)
Waddlia	*chondrophila*	Miscarriage, Respiratory infections	Unknown Unknown	Abortion (bovines)

In summary, various environmental *Chlamydiae* were discovered in the past decades revealing an unprecedented diversity of the phylum and some species, like *W. chondrophila*, might have an impact on human health. Remarkably, these bacteria are not mere phylogenetic relatives of the *Chlamydiaceae*; they also exhibit the obligate intracellular lifestyle and the unique biphasic developmental cycle.

3.5 The developmental cycle — a key feature of *Chlamydiae*

The most striking similarities of all *Chlamydia* are (I) an intracellular parasitic life-style that relies on nutrient uptake of, for instance, amino acids, vitamins, cofactors, nucleotides and sugars from the host (reviewed in McClarty, 1994); (II) a biphasic developmental cycle that is characterized by the alternation of infectious elementary bodies (EBs) and non-infectious reticulate bodies (RBs) (reviewed in Abdelrahman and Belland, 2005; Horn, 2008); and (III) the establishment of an intracellular inclusion in which the bacteria reside and that is highly modulated by the *Chlamydia* (reviewed in Fields and Hackstadt, 2002). An illustration of the chlamydial developmental cycle and the establishment of the inclusion is shown in **Figure 1**.

Bedson and Bland first described the chlamydial development within eukaryotic host cells using light microscopy (Bedson and Bland, 1932). Thereby they discovered states of different sizes; the smaller EBs and larger forms, which were called RBs later. With the advent of the electron microscope and the ultramicrotome, these observations were confirmed and extended. Two morphologically different developmental forms with a continuous gradation of intermediates between them were soon recognized (Moulder, 1966). EBs are small, rigid forms (~0.3 µm) with an electron-dense nucleoid whereas RBs are larger (~1 µm) showing a granularly structured cytoplasm with diffuse nucleic acids. The rigidity of EBs originates from the outer membrane complex (OMC), i.e. the chlamydial cell wall, which is a highly disulphide-linked proteinaceous layer protecting the EBs (Elwell, Mirrashidi and Engel, 2016).

Major components of this OMC network are the cysteine-rich proteins OmcA and OmcB, a diverse family of polymorphic membrane proteins (Pmp) and the outer membrane protein (Omp) beta-barrel porins OmpA and PorB (Liu *et al.*, 2010). Pmps resemble autotransporter proteins, which are probably unique to the *Chlamydiaceae* (Collingro *et al.*, 2011). Therefore, they have been considered to be involved in the adaptation to hosts and Pmps facilitate the initial adhesion of EBs to the host cell (Wheelhouse *et al.*, 2009; Collingro *et al.*, 2011). Interestingly, many of the proteins identified in the OMC of the *Chlamydiaceae* such as PorB, OprB, and most Pmps have no homologs in *Simkaniaceae*, *Waddliaceae*, and the *Parachlamydiaceae*, which implies differences in the regulation of endocytosis for host entry (Collingro *et al.*, 2011). In this context, a novel *ompA* family was found in *W. chondrophila*, consisting of 11 putative porins and for two of these, i.e. OmpA2 and OmpA3, adhesive properties could be demonstrated (Bertelli *et al.*, 2010; Kebbi-Beghdadi *et al.*, 2015).

Independent of the chlamydial species, EBs are extra-cellular forms, and as such responsible for dissemination by invading the host cells via endocytosis (reviewed in Abdelrahman and Belland, 2005). Beside the initial adhesion to the host cell it has been indicated that EB internalization is mediated by type III secretion system (T3S) effectors (reviewed in de Barsy and Greub, 2013). All *Chlamydia* encode a functional T3S apparatus (T3SA), which serves as a molecular syringe translocating effector proteins (T3SE) directly from the bacteria into the host cytosol (reviewed in Buttner, 2012). The chlamydial T3SA appears to be similar to other bacteria with respect to composition and thus also likely in functions (Betts-Hampikian and Fields, 2010). Given the limited metabolism of EBs, it has been indicated that the T3SA and the effectors must be prepackaged during the late stage of infection (Mueller, Plano and Fields, 2014) and that the T3SA is activated upon contact to a host cell. Strong evidence for the preloading of EBs with T3SA was provided by the discovery of the translocated actin recruiting phosphoprotein (TARP). TARP is translocated into the host cytosol upon EB attachment (Jewett *et al.*, 2010), facilitates the recruitment of actin filaments, and is the most

abundant T3SE in *C. trachomatis* EBs (Saka *et al.*, 2011). Although the detailed mechanism of chlamydial entry has yet to be elucidated, evidence suggests that TARP mediates internalization (Betts, Wolf and Fields, 2009).

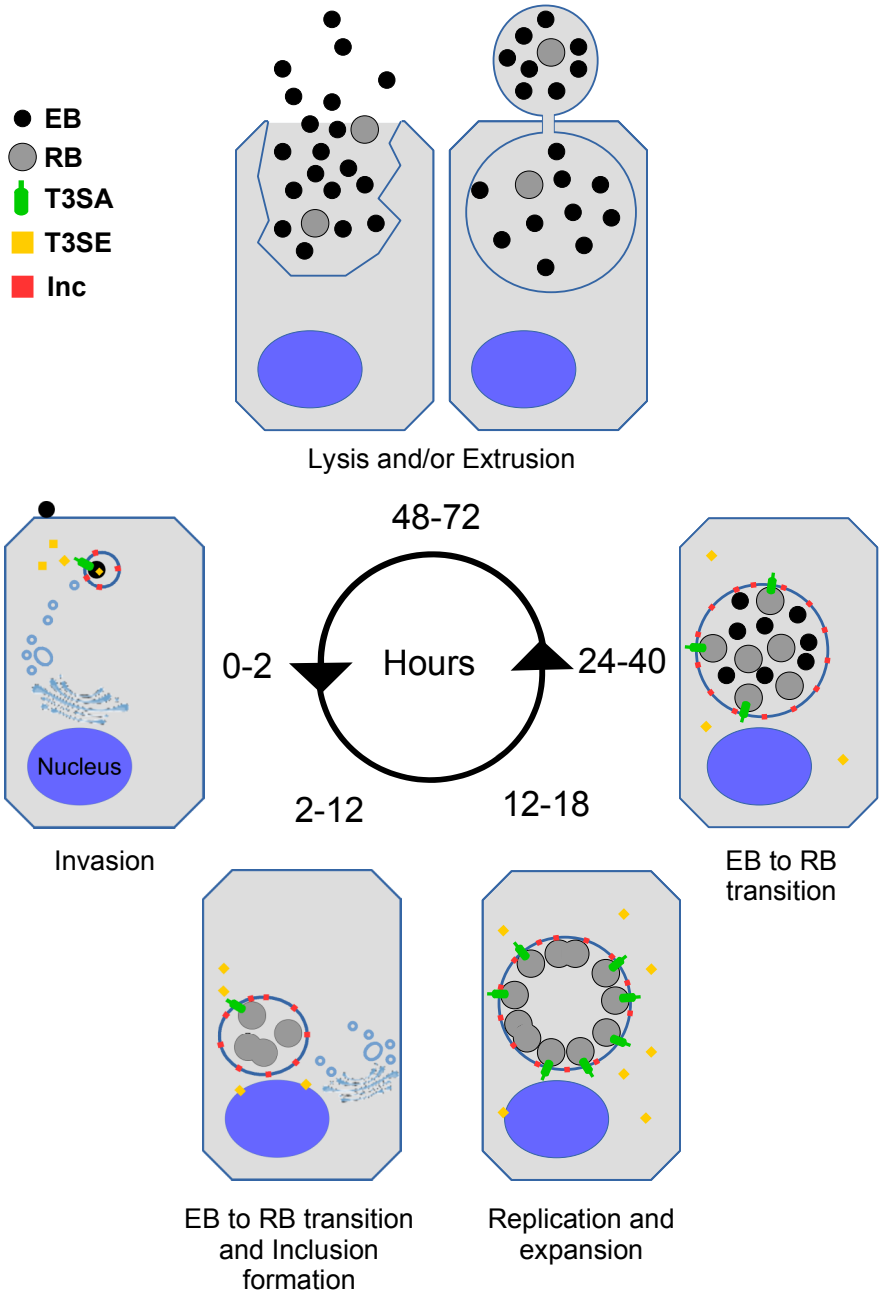


Figure 1. The developmental cycle is a key feature of all *Chlamydia* (Adapted from Bastidas *et al.*, 2013; Mueller, Plano and Fields, 2014). Upon host cell attachment elementary bodies (EBs) are internalized via endocytosis. This process is facilitated by the early activation of type III secretion system (T3S) apparatuses and the subsequent secretion of T3S effectors (T3SE). During the complete developmental cycle the vacuole like compartment in which the bacteria reside is highly

modulated by the integration of inclusion membrane proteins (Inc). Incs circumvent the fusion with lysosomes but promote incorporation of a subset of vesicles containing sphingomyelin and cholesterol from the Golgi apparatus. Between 2 h and 6 h postinfection (hpi), EBs begin to differentiate into reticulate bodies (RBs). Around 12 hpi RBs can be observed dividing by binary fission and by 18–24 hpi they peak in numbers. Increasing numbers of RBs differentiate back to EBs at the late stage and continue differentiating until lysis or extrusion occurs. About 2 – 3 days are required for life cycle completion and the duration of individual stages may vary depending on the chlamydial species and culturing conditions.

The intracellular inclusion that *Chlamydia* reside in, is highly modulated during the complete developmental cycle (**Figure 1**). Modification of the inclusion circumvents host endocytic trafficking, effectively dissociating it from late endosomes and lysosomes (Elwell, Mirrashidi and Engel, 2016). Instead, chlamydial inclusions incorporate a subset of vesicles containing lipids from the Golgi apparatus (Hackstadt *et al.*, 1996; Carabeo, Mead and Hackstadt, 2003). The molecular mechanisms that *Chlamydia* utilize to control the biogenesis of vacuoles are not fully understood yet, but are facilitated by inclusion membrane proteins (Inc). An example is the virulence factor (VF) Cap1, an Inc that was associated with capturing of lipid droplets (Saka *et al.*, 2015). While certain Incs inhibit endocytic membrane fusion events and avoid destruction, Rab- as well as SNARE-interacting Incs likely promote vesicle fusion with nutrient-rich compartments (Elwell, Mirrashidi and Engel, 2016). Hence, Incs play an important role in establishing and maintaining the intracellular niche (Mueller, Plano and Fields, 2014).

Shortly after entering the host cell, the disulphide-linked network of proteins is reduced, allowing the EBs to swell and differentiate into RBs (**Figure 1**). RBs and EBs are morphologically distinct and RBs feature a granular structured cytoplasm with diffuse, fibrillar nucleic acids. In addition, their cell wall is less rigid making them more vulnerable to physical and osmotic stress (Hatch, 1996). For a long time RBs were considered the only metabolic active forms. However, metabolic and biosynthetic activities were observed for EBs in axenic (host free) culture conditions (Omsland *et al.*, 2014). Independent of this discovery, RBs are in all *Chlamydia* the state that replicates by binary

fission. RBs exist, except for extruded inclusions (**Figure 1**), only inside the host cell and during the phase of active replication, the inclusion increases rapidly in size (Abdelrahman and Belland, 2005). Furthermore, RBs are in contact with the host cytosol via the injectisome (**Figure 1**), which secrets additional T3SE (reviewed in Mueller, Plano and Fields, 2014). After various cycles of replication RBs differentiate into EBs which are released by lysis or extrusion (**Figure 1**). Besides the classical developmental forms, *Chlamydia* can also enter into an altered growth state for persistence, which is characterized by large RB like states that cannot undergo binary fission or differentiation into EBs. Until now, several inducers have been identified to cause chlamydial persistence *in vitro*, including antibiotics, interferon-gamma (IFN-c) or nutrient depletion (reviewed in Borel *et al.*, 2014).

3.6 NGS – a technique for studying genetically inaccessible organisms

Studying the function of individual chlamydial genes has been hampered by technical challenges associated with the genetic transformation of *Chlamydia*. This lack of tools for the creation of mutant strains was primarily based on the intracellular life style. While extracellular EBs exhibit a reduced metabolism and highly compacted DNA, RBs are shielded by the host as well as the inclusion membrane. Furthermore, *Chlamydia* cannot be cultured in axenic media, thus in absence of the host. Therefore, it was not until 2009, when the first targeted mutagenesis in *Chlamydia* was reported (Binet and Maurelli, 2009). The proof of principle and the effort that was put into improving the transformation of *Chlamydia* led to various targeted transformation methods and mutant libraries (reviewed in Hooppaw and Fisher, 2016). However, despite the progress made, genetic manipulation of *Chlamydia* is still far from being a routine task.

As for a plethora of other biological fields advances in chlamydial research were highly stimulated by the development of next-generation-sequencing (NGS). These techniques comprise pyrosequencing (Roche), reversible terminator (Illumina), semiconductor (Ion Torrent) or real time sequencing (Pacific Biosciences). In general, NGS techniques

constitute various strategies that rely on a combination of template preparation, sequencing, alignment and/or assembly methods in a massive parallelized manner (reviewed in Metzker, 2010). For *Chlamydia* several comparative genome analyses have been performed yielding valuable insights into the gene content, organization and evolution of chlamydial genomes (Kalman et al., 1999; Collingro et al., 2011; Voigt, Schöfl and Saluz, 2012). In addition to the genome some transcriptome studies of *Chlamydia* have been conducted and permitted first insights regarding gene expression patterns in *C. trachomatis and C. pneumoniae*. In 2003, the first chlamydial transcriptome study was published covering the developmental cycle of *C. trachomatis* and IFN-c mediated persistence (Belland, Nelson, *et al.*, 2003). This was followed by an analysis of *C. pneumoniae* that focused again on the developmental cycle and iron-mediated persistence (Mäurer *et al.*, 2007). Both studies used microarrays, an approach by which oligonucleotides of known sequence are plotted on a surface followed by hybridization with isolated RNA. However, the complementary binding allows only detecting and quantifying of predefined sequences. With the development of RNA-Sequencing (RNA-Seq) it was possible to analyze transcriptomes without predefining detectable transcripts. RNA-Seq uses NGS methods to sequence complementary DNA (cDNA) libraries generated from total or fractionated RNA (Wang, Gerstein and Snyder, 2009). An additional advantage of RNA-seq over microarrays is the increased dynamic range (up to 5 log-folds) of transcript detection and a lower background level (Agarwal *et al.*, 2010). Because of these advantages a plethora of studies using RNA-Seq were published in the past decades. The first RNA-Seq analysis of *Chlamydia* was already published in 2010. Albrecht *et al.* isolated EBs and RBs of *C. trachomatis* L2b in order to apply RNA-Seq of both developmental stages (Albrecht *et al.*, 2010). In addition, transcriptional start sites (TSSs) of genes were annotated and 43 non-coding RNAs were found. In a follow-up study a similar approach was conducted for *C. pneumoniae* CWL-029 (Albrecht *et al.*, 2011). Moreover, other approaches like a dual RNA-Seq study

investigating the *C. trachomatis* and the corresponding host cell transcriptome were published (Humphrys *et al.*, 2013).

In summary, genome and transcriptome sequencing were extremely important avenues of chlamydial research. The following comparisons yielded valuable insights into the gene content, organization and evolution of chlamydial genomes and some gene expression studies on the transcript level have been performed.

3.7 The gene content of *Chlamydiaceae* species is similar but host preference and virulence phenotype differ

The genomes of all *Chlamydiaceae* are characterized by a reduced size of about 1.1 megabase pairs (Mbp) containing only ~1,000 coding DNA sequences (CDSs), few degraded open reading frames (ORFs) and a high coding density (~90%) indicating a highly optimized usage of the available coding capacity (Voigt, Schöfl and Saluz, 2012). Gene content and order is highly conserved among the chlamydial genomes and comparison of e.g. *C. psittaci*, *C. abortus*, *C. caviae* and *C. felis* revealed a high degree of consistency, i.e. 872 (81%) of a total of 1,074 CDSs are shared among the species (Voigt, Schöfl and Saluz, 2012). However, despite the close resemblances on the genome level host specificity and virulence among these pathogens differ substantially (**Table 1**). This feature of chlamydial species culminates in the observation that individual isolates of the same species differ in their phenotypes and growth rates (Last *et al.*, 2018; Islam *et al.*, 2019).

The most striking example of phenotypic diversity despite genomic similarity on the species level is seen for the close relatives *C. psittaci* and *C. abortus* that share 915 (93%) of their protein coding genes while infecting different hosts and tissues (**Table 1**). This is even more pronounced when one considers that *C. abortus* encodes merely 933 CDSs and of the 18 genes not shared with *C. psittaci* only 8 are unique to *C. abortus*. In turn, *C. psittaci* encodes also only 13 unique CDSs (Voigt, Schöfl and Saluz, 2012). While the proteins, that are not conserved among species, are often thought to represent

species-specific functions in niche adaptions (Dean, Meyers and Read, 2006), in fact, all the better investigated VFs (e.g. OmpA, OmcB or IncA) have well-conserved orthologs. With few exceptions, all the species-specific genes encode short hypothetical proteins with no significant homology to any gene model of known function (Voigt, Schöfl and Saluz, 2012). Therefore, there is little to suggest that the different host affinities and virulence phenotypes of different chlamydial species can be easily correlated with the presence or absence of different sets of genes. In contrast, marked differences in the capability to colonize host tissue were shown to correlate with the expression rates of a number of homologous factors related to virulence in *C. psittaci* and *C. abortus* (Braukmann *et al.*, 2012).

On the other side single nucleotide polymorphisms of conserved genes in chlamydial isolates are considered to be involved in host preference, disease severity and tissue tropism. The best studied representative is *C. trachomatis* of which the two biovars have less than 1% genetic differences (Harris *et al.*, 2012; Seth-Smith *et al.*, 2013). In addition, for all chlamydial species with genomes sequenced from multiple isolates polymorphisms were found and from the beginning suggested being involved in conferring isolate specific traits (Read *et al.*, 2013; Weinmaier *et al.*, 2015; Hadfield *et al.*, 2017; Seth-Smith *et al.*, 2017). However, for example polymorphisms in the surface-exposed protein *ompA* have been repeatedly postulated but failed to actually reveal pathobiological distinctions between the various *C. trachomatis* isolates (Byrne, 2010). Actually it was not until 2018, when a first association between polymorphisms in *pmpD* and *tarP* with *C. trachomatis* ocular localization as well as *yjfH, glgA* and *pmpE* polymorphisms associated with disease severity were found (Last *et al.*, 2018).

3.8 The *Waddlia chondrophila* genome

The most apparent genomic difference of environmental *Chlamydia* like *Waddliaceae*, *Simkaniaceae*, and *Parachlamydiaceae* as compared to the *Chlamydiaceae* is the 2- to 3-fold larger genome size of the former (Horn *et al.*, 2004; Bertelli *et al.*, 2010; Collingro *et*

al., 2011). However, despite the larger genomes and higher coding capacity (~2,000 CDSs), a similar host dependence and obligate intracellular life style has been observed (Horn *et al.*, 2004). Interestingly, an analysis showed that 560 CDSs were conserved among chlamydial families representing a core genome, which might set the characteristics shared among all *Chlamydia* such as the unique developmental cycle or the establishment of the intracellular inclusion (Collingro *et al.*, 2011). The complete genome sequence of *W. chondrophila* WSU 86-1044 was sequenced in 2010 (Bertelli *et al.*, 2010) and the draft genome published of *W. chondrophila* 2032/99 in 2011 (Collingro *et al.*, 2011). Both isolates possess one chromosome of ~2.1 Mbs with a GC content of 43-44% (Bertelli *et al.*, 2010; Collingro *et al.*, 2011). The *W. chondrophila* WSU 86-1044 genome displays two sets of ribosomal RNA (rRNA) operons, 37 transfer RNA (tRNA) genes and 1,934 CDSs. It also encodes several homologous VFs shared with the *Chlamydiaceae* and the functional T3SA (Collingro *et al.*, 2011). However, important differences are that *W. chondrophila* possesses improved capabilities for the *de novo* synthesis of amino acids, lipids, nucleotides and co-factors (Bertelli *et al.*, 2010). These features and the capability of an independent energy production by oxidative phosphorylation (Bertelli *et al.*, 2010) make *W. chondrophila* a promising candidate for cultivation in axenic media, which would allow an unprecedented number of assays to study e.g. the chlamydial biology or therapeutic approaches.

3.9 Differential RNA-Seq and TSS annotation

The two chlamydial RNA-Seq studies of Albrecht *et al.* used a technique that increases the information derived from transcriptome data, called differential RNA-Seq (dRNA-Seq). The dRNA-Seq approach is based on the sequencing of a library selectively enriched for primary transcripts and comparing this to a library that was not enriched (Sharma *et al.*, 2010; Sharma and Vogel, 2014). Principally, the enrichment of primary transcripts can be achieved because these carry a 5’-triphosphate end (5’PPP), whereas processed transcripts including the abundant rRNAs or tRNAs carry a 5’-monophosphate (5’P) or,

less frequently, a 5'-hydroxyl group. The enrichment of primary transcripts is accomplished by treatment of the RNA with an enzyme called Terminator 5'-Phosphate-Dependent Exonuclease (TEX), which digest RNAs harboring a 5'P end (Sharma and Vogel, 2014). Subsequently, a test for differential gene expression in +TEX and -TEX libraries allows the detection of primary transcripts, which are enriched in the +TEX library. Moreover, dRNA-Seq can be used to improve genome annotations by mapping TSSs at single nucleotide resolution (**Figure 2**). This allows the identification of 5' untranslated regions (5'UTRs), promoter motifs, unknown ORFs, antisense (asRNAs) or small non-coding RNAs (sRNAs). Since 2010 several studies using dRNA-Seq to detect primary transcripts and annotate TSSs have been published for various bacteria including *Helicobacter pylori*, *Salmonella enterica*, *Propionibacterium acnes*, *Escherichia* (*E.*) *coli* (Sharma *et al.*, 2010; Kröger *et al.*, 2012; Lin *et al.*, 2013; Thomason *et al.*, 2015) and the chlamydial species *C. trachomatis* and *C. pneumoniae* (Albrecht *et al.*, 2010, 2011).

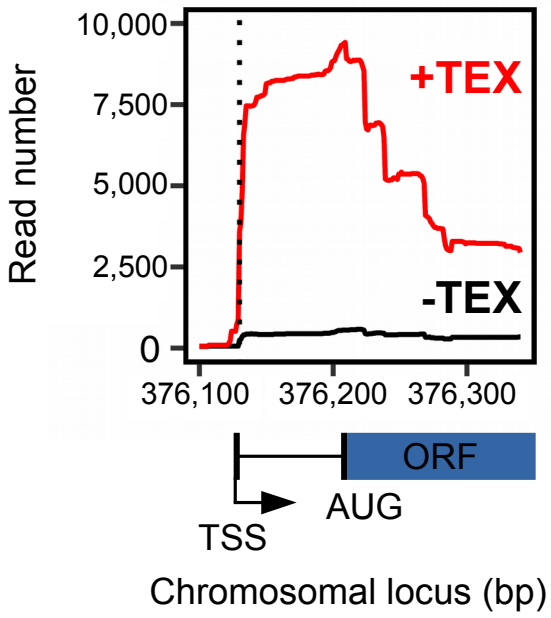


Figure 2. A dRNA-Seq specific cDNA enrichment pattern can be observed at the primary 5'-ends of genes. Treatment with TEX (+TEX) enriches cDNAs towards the nuclease-protected 5'-end exemplified here for the *C. psittaci* gene CPS0B_RS01660. Subsequently, a comparison with the untreated library (-TEX) allows the annotation of the TSS (dotted line). Knowing the exact position of TSS the 5'UTR as well as promoter sequences can be analyzed.

To note, treatment of the RNA with TEX allows the exact determination of TSSs at single nucleotide resolution (**Figure 2**) as previously demonstrated for various bacteria. However, the studies often used visual inspection of the dRNA-Seq data for TSS annotation, which is tedious considering that the complete genome has to be manually

examined for differences in the +TEX and -TEX libraries. Therefore, the application of TSS prediction tools developed specifically for dRNA-Seq data like TSSpredator (Dugar *et al.*, 2013), TSSAR (Amman *et al.*, 2014) and TSSer (Jorjani and Zavolan, 2014), should be much faster. However, application of e.g. TSSpredator on *E. coli* dRNA-Seq data lead to 14,868 TSSs predictions (Thomason *et al.*, 2015) for an organism encoding merely ~4,200 genes. About 3.5 times more predicted TSSs would imply 3.5 times individual more transcripts than annotated genes. This number appeared too high even thought RNA-Seq revealed that the transcriptomes of bacteria are much more diverse than previously expected (reviewed in Sorek and Cossart, 2010; Georg and Hess, 2011). Therefore, the large numbers of TSS annotated might be due to non-optimal parameter settings allowing to many false positives. The problem of potential false positive TSS annotations derived from high throughput data becomes even clearer when considering that another study using Cappable-seq for TSS annotation in *E. coli* annotated 16,953 TSSs (Ettwiller *et al.*, 2016). Of these, 5,602 were also found by the dRNA-Seq approach, which would imply 22,986 TSSs found in *E. coli* by both techniques together. Consequently, this would mean the presence of 5.5 times more individual transcripts than annotated genes. For this reason, there is a need to optimize the parameters for TSS annotation, which might substantially improve the reliability of predictions.

3.10 Aims of the study

So far transcriptome analyses for the major human pathogens *C. trachomatis* and *C. pneumoniae* have been conducted focusing on gene expression during development (Belland, Zhong, *et al.*, 2003; Mäurer *et al.*, 2007) and also purified EBs and RBs (Albrecht *et al.*, 2010, 2011). However, comparative studies between species like they were performed on the genome level for *Chlamydia* (Kalman et al., 1999; Collingro et al., 2011; Voigt, Schöfl and Saluz, 2012) have not yet been carried out. Only such a comparison would allow the detection of transcriptional differences in the expression of homologous genes. Previous work from 2012, comparing *C. psittaci* and *C. abortus*

infections in chicken embryos suggested that relative differences in the expression levels of bacterial VFs rather than differences in gene content *per se* play a major role in determining differences in the invasiveness and propagation of *Chlamydia* (Braukmann *et al.*, 2012). Therefore, in this work dRNA-Seq was applied to purified EBs and RBs of *C. psittaci*, *C. abortus* and *W. chondrophila*. Using a comparative approach, the transcriptomes of the infectious and non-infectious states as well as the expression of homologous VFs and immunogenic factors were analyzed. The rationale was that *C. psittaci* and *C. abortus* have highly similar genomes and a well-documented zoonotic potential but differ substantially in disease outcome and host preference (Longbottom and Coulter, 2003; Horn, 2008; Voigt, Schöfl and Saluz, 2012). It is reasonable to assume that these differences originate from the presence of species-specific genes and polymorphisms, but may also be due to the differential expression of homologous VFs (Braukmann *et al.*, 2012). The more distant relative *W. chondrophila* was included as a third model because infections occur in the human respiratory tract and *W. chondrophila* might be associated with abortion (Haider *et al.*, 2008; Baud, Goy, *et al.*, 2014). This resembles tissue tropism and disease outcomes of *C. psittaci* and *C. abortus*.

Secondly, this work focused on the prediction of TSSs, which was motivated by the hypothesis that a comparative analysis like their relative position to ORFs as well as derived 5'UTR and promoter sequences, should provide valuable information about gene regulatory mechanisms in *Chlamydia*. Therefore, this work focused on both, the optimization of parameters and combination of TSS annotation tools that use dRNA-Seq data. This approach should principally result in more precise annotations.

Overall, this work focused on three major goals: (I) intra-species comparison of gene expression in EBs and RBs of *C. psittaci*, *C. abortus* and *W. chondrophila* (II) inter-species comparison of gene expression, especially of VFs and (III) reliable annotation of TSSs to analyze the corresponding 5'UTRs and promoters as well as their conservation among *Chlamydia*.

4 Materials and Methods

The titer determination, purification of EBs and RBs, RNA isolation and DNase I digest of *W. chondrophila* were performed by Mareike Scheven, a Master student at the CMB department till September 2014, which I supervised. The materials and methods for these works were adapted from (Scheven, 2014) and included in this thesis for completeness. Purification of *C. psittaci* EB and RB fractions for electron microscopy was carried out by Dominique Präzsch, a Master student at the CMB department till April 2015, which I supervised (Praetzsch, 2015).

4.1 Materials

4.1.1 Instruments and technical equipment

Table 2. List of instruments and technical equipment.

Instrument	Manufacturer
2100 Bioanalyzer Instrument set	Agilent Technologies (Waldbronn, Germany)
Balance	Sartorius (Goettingen, Germany)
Beckman Cordless Tube Topper™	Beckman Coulter GmbH (Krefeld, Germany)
Beckman Optima L-70 Ultracentrifuge	Beckman Coulter GmbH (Krefeld, Germany)
Biofuge fresco	Heraeus GmbH (Haunau, Germany)
Eppendorf Centrifuge 5417R	Eppendorf AG (Hamburg, Germany)
Eppendorf Centrifuge 5702R	Eppendorf AG (Hamburg, Germany)
Gel Doc™ EZ Imager	Bio-Rad (Munich, Germany)
HERAcell 150 CO_2 Incubator	Fisher Scientific GmbH (Schwerte, Germany)
HERAcell 150i CO_2 Incubator	Fisher Scientific GmbH (Schwerte, Germany)
Homogenizer precellys® 24	PEQLAB Biotechnologie GmbH (Erlangen,Germany)
IKAMAG® RCT magnetic hot plate stirrer	IKA® -Werke GmbH & Co.KG (Staufen, Germany)
IMPLEN NanoPhotometer®	Implen GmbH (Munich, Germany)
Incubation/Inactivation Water Bath 1008	GFL (Burgwedel, Germany)
Laminar Box HERAsafe® KS 12	Thermo Electron LED GmbH (Langenselbold, Germany)
Light microscope Axiovert 25	Carl Zeiss MicroImaging (Gottingen, Germany)
Magnetic stand (6-tube)	Life Technologies GmbH (Darmstadt, Germany)
Microscope camera DP50	Olympus (Hamburg, Germany)
Microwave	Siemens AG (Munich, Germany)
NALGENE® Mr. Frosty™ Cryo 1 °C Freezing Container	Fisher Scientific GmbH (Schwerte, Germany)

Table 2. Continued.

Instrument	Manufacturer
NanoDrop™ ND-1000 Spectrophotometer	PEQLAB Biotechnologie GmbH (Erlangen, Germany)
Olympus BX-51M microscope	Olympus (Hamburg, Germany)
Peristaltic pump type MS-CA4	ISMATEC (Wertheim, Germany)
Pharmacia Electrophoresis Power Supply – EPS 600	GE Healthcare Life Sciences (Amersham, United Kingdom)
Pipetboy accu-jet® pro	BRAND GmbH & Co. KG (Wertheim, Germany)
Power Pac™ HC	Bio-Rad (Munich, Germany)
Rotanta 460 RS centrifuge	Andreas Hettich GmbH & Co.KG (Tuttlingen, Germany)
Safe 2020.9 Class II Biological Safety Cabinet	Thermo Electron LED GmbH (Langenselbold, Germany)
SevenEasy InLab® Semi-Micro pH meter	METTLER TOLEDO GmbH (Gießen, Germany)
StepOnePlus™ Real-Time PCR System	Life Technologies GmbH (Darmstadt, Germany)
SW 28 Swinging-Bucket Rotor	Beckman Coulter GmbH (Krefeld, Germany)
TC10™ automated cell counter	Bio-Rad (Munich, Germany)
Thermomixer® comfort	Eppendorf AG (Hamburg, Germany)
UNIVERSAL 32 R centrifuge	Andreas Hettich GmbH & Co.KG (Tuttlingen, Germany)
Veriti™ 96 Well Fast Thermal Cycler	Life Technologies GmbH (Darmstadt, Germany)
Vortex Genie 2™	Bender & Hobein GmbH (Zurich, Switzerland)
Wide Mini-SUB® Cell GT	Bio-Rad (Munich, Germany)

4.1.2 Consumables

Table 3. List of plastic ware and consumables.

Consumable	Manufacturer
2 ml Soft tissue homogenizing CK14 tubes	PEQLAB Biotechnologie GmbH (Erlangen, Germany)
Agencourt® AMPure® XP magnetic beads	Beckman Coulter GmbH (Krefeld, Germany)
BD FalconTM Serological Pipets (1 ml, 2 ml, 5 ml, 10 ml, 25 ml)	BD Biosciences (Heidelberg, Germany)
BD FalconTM Tubes (15 ml, 50 ml)	BD Biosciences (Heidelberg, Germany)
Cell culture flasks (25 cm^2 , 75 cm^2 , 175 cm^2)	Greiner Bio-One GmbH (Frickenhausen, Germany)
Cell scraper (24 cm)	TPP Techno Plastic Products AG (Trasadingen, Switzerland)
Counting Slides	Bio-Rad (Munich, Germany)
Cryotubes	SARSTEDT AG & Co. (Nuembrecht, Germany)
DIFFINITY RapidTip® 2	Diffinity Genomics, Inc. (West Henrietta, USA)
Disposable Pasteur Pipette	ELKAY ÉIREANN (Costelloe, Ireland)
Fast 96-Well Reaction Plate (0.1 ml)	Life Technologies GmbH (Darmstadt, Germany)
Glas Pasteur pipette	Carl Roth GmbH & Co. KG (Karlsruhe, Germany)
MicroAmp® Clear Adhesive Film	Life Technologies GmbH (Darmstadt, Germany)
MicroAmp® Optical Adhesive Film	Life Technologies GmbH (Darmstadt, Germany)
Microscope slides and coverslips	Carl Roth GmbH & Co. KG (Karlsruhe, Germany)
Minisart syringe filter (0.2 µm, hydrophilic)	SartoriusStedim Biotech GmbH (Goettingen, Germany)
NEOJECT® disposable cannula (21G x 1½'', 0.8 x 40 mm)	Dispomed Witt oHG (Gelnhausen, Germany)

Table 3. Continued.

Consumable	Manufacturer
Omnifix® disposable syringe (10 ml, 50 ml)	B. Braun Melsungen AG (Melsungen, Germany)
Parafilm M®	Pechiney Plastic Packaging (Chicago, USA)
Pipette filter tips TipOne® (10 µl, 100 µl, 1000 µl)	STARLAB GmbH (Hamburg, Germany)
Pipette tips (10 µl, 100 µl, 1000 µl)	SARSTEDT AG & Co. (Nuembrecht, Germany)
Polyallomer Bell-Top Quick-Seal® Centrifuge Tubes, 33 ml	Beckman Coulter GmbH (Krefeld, Germany)
Safe-Lock Tubes (1.5 ml, 2.0 ml)	Eppendorf AG (Hamburg, Germany)
Sartolab RF 250 Filter System (0.22 micron PES, 250 ml)	Sartorius Stedim Biotech GmbH (Goettingen, Germany)

4.1.3 Chemicals

If not specified, chemicals were purchased from Bioline (Luckenwalde, Germany), Carl Roth GmbH (Karlsruhe, Germany), Fisher Scientific GmbH (Schwerte, Germany), Life Technologies GmbH (Darmstadt, Germany) MERCK KGaA (Darmstadt, Germany), Roche (Mannheim, Germany), Sigma-Aldrich (Steinheim/Seelze, Germany) or VWR (Darmstadt, Germany).

4.1.4 Cell line

The human cell line HEp-2 (CCL-23) was obtained from the American Type Culture Collection. HEp-2 is an immortal human epidermoid carcinoma cell line in which infections with *Chlamydia* resulted in high yields (Li *et al.*, 2005).

4.1.5 Bacterial strains

C. psittaci, *C. abortus* and *W. chondrophila* strains were kindly provided by Dr. Konrad Sachse (Friedrich-Loeffler-Institute, Jena, Germany). The *C. psittaci* strain 02DC15 was initially isolated from an aborted calf fetus in 2002 (Sachse *et al.*, 2009); the *C. abortus* S26/3 was isolated from a vaccinated ewe that aborted in 1979 (McClenaghan, Herring and Aitken, 1984); and *W. chondrophila* 2032/99 was isolated from aborted bovine fetus (Henning *et al.*, 2002).

4.1.6 HEp-2 culture medium

Dulbeccos's Modified Eagle Medium (DMEM) with Glucose (4.5 g/l), L-glutamine, sodium pyruvate (PAA Laboratories) was used as culture medium with supplementation of 10% v/v fetal bovine serum (FBS).

4.1.7 *Chlamydia* transport medium

Sucrose-Phosphate-Glutamate transport medium (SPG) was prepared according to (Bovarnick, Miller and Snyder, 1950). Additionally, 1% w/v bovine serum albumin was added. Filter before use, store at -20°C.

4.1.8 Buffers and reagents

All buffers and reagents were prepared with ultrapure water.

0.1 M sodium phosphate buffer pH 7.5
7 ml 0.2 M NaH_2PO_4 x H_2O
43 ml 0.2 M Na_2HPO_4 x 7 H_2O
50 ml water
prepare fresh before use

0.5% (w/v) citric acid solution (prepare fresh before use)

0.8% (w/v) aqueous malachite green oxalate; filter before use, store at room temperature (RT)

10x Bionic buffer
100 nM NaOH
adjust pH to 8.0 with boric acid
store at RT

Basic fuchsin solution (prepare fresh and filter before use)
42 ml 0.1 M sodium-phosphate buffer pH 7.5
4 ml basic fuchsin stock solution

Basic fuchsin stock solution
2.5 ml [1% (v/v)] phenol
25 ml [10% (v/v)] 99.5% ethanol
2.5 g [1% (w/v)] fuchsin basic
225 ml water
incubate 48 h at 37°C and filter before use, store at RT

4.1.9 Kits

Table 4. List of kits.

Kit	Manufacturer
Agilent High Sensitivity DNA Kit	Agilent Technologies (Waldbronn, Germany)
Agilent RNA 6000 Pico Kit	Agilent Technologies (Waldbronn, Germany)
RNA Clean & Concentrator™-5	ZymoResearch (Irvine, USA)
ScriptSeq™ Complete Gold Kit (Epidemiology)	Illumina (San Diego, USA)

4.1.10 Enzymes and inhibitors

Table 5. List of enzymes and inhibitors.

Enzyme	Manufacturer
DNase I, RNase free (1 U/µl)	Fisher Scientific GmbH (Schwerte, Germany)
iScript™ Reverse Transcription Supermix	Bio-Rad (Munich, Germany)
RNase A, DNase and protease-free (10 mg/ml)	Fisher Scientific GmbH (Schwerte, Germany)
SsoFast™ EvaGreen® Supermix With Low ROX	Bio-Rad Laboratories GmbH (Munich, Germany)
SUPERase-In™ (20 U/µl)	Life Technologies GmbH (Darmstadt, Germany)
Terminator™ 5´-Phosphate-Dependent Exonuclease (1 U/µl)	(Epicentre, USA)
Trypsin	MERCK KGaA (Darmstadt, Germany)

4.2 Primers

Table 6. List of primers.

Name	Primer sequence (5' → 3')
Table 6.1. Control of DNaseI digestion by RT-qPCR	
16S_rRNA-fw[a]	ACCCTAAGTGTTGGCAACTAAC
16S_rRNA-rv[a]	TTTCCGCAAGGACAGATACAC
ompA-for	CGTTGCACCAACAGCTACAC
ompA-rev	GCACTATGTGGGAAGGTGCT
recA-fw[b]	AGGTCGAGCGCTGAAATTCT
recA-rv[b]	AATTCCGCAGCTCGAAAAG
Table 6.2. Primers for homologs among *C. psittaci* and *C. abortus**	
groEL-forward (fw)	TTGCCCGAGCGTGTATCTTT
groEL-reverse (rv)	TTCTTCGCGTACAGGGACTA
pmpA-fw	GGACTCTGTCCATTACAGGTGG
pmpA-rv	CCTGAGCGGATCTCAGCA
cap1-fw	TCAAAGAATTGCTTCGCACAC
cap1-rv	TGATCGCCCTTAGTCAGGTAG
sinC-fw	GCTTCAGGGCCAAAGCGTC
sinC-rv	TGGAGGTTGAGGAGCTTTACG
copB_1-fw	AACGCTGAAATGCAAAGCGA
copB_1-rv	GCTGCAGCTTCTGCGGAT
incA-fw	AACATGGTGGATGCTGTGAG
incA-rv	ACGCTGTTCTAAGTAAAGTAAATGC
ompA-fw	TCTTGCAAATTGCTTCGATTCAGA
ompA-rv	GATTAAGCGTGCTTCACCAGT

Table 6. Continued.

Table 6.3. Primers for homologs among *C. psittaci*, *C. abortus* and *W. chondrophila***	
hrtA_psiab-fw	GCGGATAGCAGAAGTCCCTC
hrtA_psiab-rv	ACAGGACGGTTACACTTAGGA
hrtA_wch-fw	ACAGCCGATCGACAACAACT
hrtA_wch-rv	AGCATTGCGTAAGCTGGAGA
mip_psiab-fw	GCGCACCTTTAACTGAATCCG
mip_psiab-rv	TGTACTTCTACGACACCTGCAT
mip_wch-fw	GAGAGTTGGAGAGACGGCAC
mip_wch-rv	GGCTAACCGACGTTTTTCCAT
hctA_psiab-fw	CGTGACTTAGACAAGGCCGA
hctA_psiab-rv	GCAGGCTTGCGAGTCAT
hctA_wch-fw	CTGCTAAGAAAGCCGTTGCC
hctA_wch-rv	AGCTCTAGCTCTAGCTGTCG
dnaK_psiab-fw	CCGTAATGGAAGGTGGGCAAG
dnaK_psiab-rv	CCGATGAAACGTTTTGTAGAAGC
dnaK_wch-fw	GGGTCAATCCGGACGAAGTT
dnaK_wch-rv	TCGGCTGCAGTGGAAAAGAT
omcB_psiab-fw	GGACAACGCGTTCTTTCCTT
omcB_psiab-rv	AGACCAGTCAGCTCCAGAGA
omcB_wch-fw	AACGTTCGCGTTGTTGTGAG
omcB_wch-rv	ACGTGGGCCAAAGTTATCGT

a) Binet and Maurelli, 2005

b) Barkallah *et al.*, 2013

*Primers were designed to the corresponding consensus sequences.

**Because of sequence differences additional primers were designed for *W. chondrophila*.

4.3 Methods

4.3.1 Cell line cultivation

HEp-2 cultivation was carried out in DMEM (+10% FBS) at 37°C in 5% CO_2 and 95% humidity (standard conditions for all experiments). Growing of cells was controlled by light microscopy and at 90% confluency the cells were passaged. Therefore, cells were washed with 1x PBS and incubated with 0.05% trypsin/0.02% EDTA (w/v) solution until cells detached. The cells were collected in DMEM and centrifuged at 500 RCF at RT for 3 min. The supernatant was aspirated and the cell pellet was resuspended in DMEM. Passaging was performed in a 1:4 – 1:6 ratio three times a week.

4.3.2 Chlamydial in-house stock preparation

A preculture of HEp-2 cells was infected with the *C. psittaci*, *C. abortus* and *W. chondrophila* stocks. Therefore, 5 µl of the chlamydial stock were added to each of eight cell culture flasks (25 cm^2), which contained HEp-2 cells at 90% confluency, followed by centrifugation at 1,800 RCF at 37°C for 1 h. After 48 h incubation at 37°C, the medium was aspirated and cells were collected by scrapping and pooled in 5 ml DMEM. 5 µl of preculture were added to each of 20 cell culture flasks (25 cm^2), followed by centrifugation at 1,800 RCF at 37°C for 1 h. After 48 h of incubation at 37°C, the medium was removed from the main culture flasks. Infected cells were collected by scrapping, pooled in 15 ml SPGA buffer, homogenized thoroughly and aliquoted in cyrotubes (0.5 ml each). The aliquots were slowly frozen to -80°C using a NALGENE® Mr. Frosty™ Cryo 1°C Freezing Container and stored at -80°C until use.

4.3.3 Immunohistochemical staining of *C. psittaci* and *C. abortus*

For a defined multiplicity of infection (MOI) of HEp-2 cells it was essential to determine the number of inclusion forming units per ml (IFU/ml) of the *C. psittaci*, *C. abortus* and *W. chondrophila* in-house stocks. Thereby, the MOI expresses the ratio of the total number of infectious *Chlamydia* inoculated to the total number of host cells (Friis, 1972). For IFU/ml determination, four cell culture tubes containing a glass slip for microscopy were seeded with $2x10^5$ HEp-2 cells in 1 ml DMEM supplied with 10% FBS and incubated at 37°C for 24 h. Three cell culture tubes were infected with 200 µl of a 10^{-4} dilution of the chlamydial in-house stock in PBS. In addition, 2 cell culture tubes were mock infected with 200 µl culture media. The tubes were centrifuged at 1,800 RCF 37°C for 1 h and afterwards the supernatant was replaced with fresh DMEM supplied with 10% FBS. After 24 h of incubation the medium was removed, cover slips were washed with methanol and 3 ml fresh methanol were added. Tubes were stored at 8°C for 30 min, then the cover slip was removed and affixed with Entellan (Merck Chemicals GmbH) on a glass slide. 3 µl IMAGEN™ reagent were mixed with 17 µl PBS and layered over the cover slip. The

reagent contains fluorescein isothiocyanate linked monoclonal antibodies against the chlamydial lipopolysaccharide. The slides were incubated in a humidity chamber at 37°C for 30 min and briefly washed with PBS and distilled water before mounting.

4.3.4 Giménez staining protocol

Infection of HEp-2 cells in cell culture tubes with *W*. *chondrophila* was carried as described in **4.3.3**. Because the antibody used in the IMAGEN™ reagent does not bind to the *W*. *chondrophila* lipopolysaccharide the inclusions were Giménez stained in order to define the IFU/ml. In contrast to the original Giménez staining protocol (Giménez, 1964), basic fuchsin working solution was prepared in a 1:11.5 dilution ratio of basic fuchsin stock solution to sodium phosphate buffer and a destaining step using 0.5% (w/v) citric acid solution was introduced after staining with basic fuchsin. The slides were then covered with 0.8% (w/v) aqueous malachite green oxalate for 30 s, washed in tap water, covered with malachite green oxalate for additional 30 s, again washed in tap water and air-dried before mounting.

The staining protocols described in **4.3.3** and **4.3.4** were used for titer determination of *C*. *psittaci*, *C*. *abortus* and *W*. *chondrophila* in-house stocks, observation of life cycle within HEp-2 cells as well as reinfection assays with purified RBs and EBs.

4.3.5 Determining the chlamydial titer

C. *psittaci*, *C*. *abortus* and *W*. *chondrophila* infected HEp-2 cells on cover slips and corresponding controls were prepared in triplicate as described in sections **4.3.3** and **4.3.4**. Slides were examined with the Fluorescence microscope BX-51 M (Olympus) at 400x magnification and pictures were taken using the Cell^D Life Science documentation software (Olympus). For titer calculation, 20 visual fields of each sample were photographed randomly. Inclusions were counted using ImageJ 1.48v and the average number was multiplied by the area factor (1,771.5) as well as the dilution factors to calculate the IFU/ml of chlamydial in-house stocks.

4.3.6 Infection of HEp-2 cells with *Chlamydia*

C. abortus and *C. psittaci* infections of HEp-2 cells (25 cm^2) were carried out with a multiplicity of infection (MOI) of 5.0, and *W. chondrophila* infection with a MOI of 0.5. The differing MOI was applied due to the devastating effects of higher *W. chondrophila* MOIs on the cell layer. Infected cells were centrifuged at 1,800 RCF and 37°C for 1 h and incubated under standard conditions for 1 h followed by medium replacement. Centrifugation of infected cells followed by media exchange was performed to synchronize the infection. Infected cells were incubated under standard conditions until harvesting for RB and EB purification. In order to isolate only characteristic states of EBs and RBs (exclusion of transition states of RBs to EBs and *vice versa*, which are called intermediate bodies), the chlamydial developmental cycle was monitored. Time points for RB isolation were selected during the exponential growth phase when the inclusions were in the range of 4–6 µm. Generally, inclusion size among the species is similar and expansion is linked to bacterial replication (Tietzel, El-Haibi and Carabeo, 2009; Nguyen *et al.*, 2011; Engstrom *et al.*, 2013; Volceanov *et al.*, 2014). RBs of *C. psittaci* were collected at 24 hpi; *C. abortus*, 36 hpi and *W. chondrophila*, 18 hpi. Time points for EB purification were selected just before host cell rupture and release of *Chlamydia* in order to collect the EBs within the first propagation cycle for all species. Thus, EBs were purified as follows: *C. psittaci*, 38 hpi; *C. abortus*, 48 hpi; *W. chondrophila*, 30 hpi. Infected cells of four (*C. psittaci*) or three (*C. abortus* and *W. chondrophila*) biological replicates were collected and stored at -80°C until use for purification of RBs or EBs.

4.3.7 Purification of chlamydial RBs and EBs

All steps for EB and RB purification were carried out on ice, using ice-cold buffers and precooled centrifuges (4°C). Infected cells were homogenized using ceramic beads in a precellys® 24 homogenizer at 5000 rpm for 10 s. Cell debris were pelleted by centrifugation at 1,000 RCF for 10 min and *Chlamyida* were collected at 10,000 RCF for 10 min. The obtained pellet was rinsed and resuspended in 2 ml SPG buffer

supplemented with Tris-HCl (10 mM, pH 7.5, RT), $MgCl_2$ (100 µM), $CaCl_2$ (50 µM), 100 µl DNase I and 4 µl RNase A. The crude extract was incubated at 37°C for 30 min and after homogenization by pipetting through a 0.8 mm gauge needle layered on top of a discontinuous Visipaque® 320 (GE Health Care, Amersham, United Kingdom) gradient. The gradient was established in a 33 ml Polyallomer Bell-Top Quick-Seal® centrifugation tube using a peristaltic pump and consisted of following layers: 2 ml crude extract, 12 ml 30% (v/v), 14 ml 37% (v/v) and 5 ml 50% (v/v) Visipaque® diluted in PBS. RBs and EBs were separated by centrifugation at 53,000 RCF in SW 28 swinging-bucket rotor in a Beckman Optima L-70 ultracentrifuge for 1 h. RBs accumulated at 37% (v/v) Visipaque® and EBs at 50% (v/v). RB and EB phases were carefully collected using the peristaltic pump, washed with PBS and centrifuged at 10,000 RCF for 10 min to pellet the RBs and EBs.

4.3.8 Quality control of RB and EB isolates by TEM

For quality control pellets were fixed overnight in 1 ml 0.1 M sodium cacodylate buffer with 2.5% (v/v) glutaraldehyde (provided by PD Dr. Martin Westermann). After fixation, pellets were washed three times in 0.1 M cacodylate buffer for 5 min at RT. Ultrathin sections, mounting and transmission electron microscope (TEM) imaging was performed by PD Dr. Martin Westermann (Jena University Hospital, Electron Microscopy Center, Jena, Germany).

4.3.9 Infection assays with RB and EB fractions

The purity of each isolated RB and EB fraction was controlled via an infection assay. EBs are infectious while RBs are non-infectious (reviewed in Abdelrahman and Belland, 2005; de Barsy and Greub, 2013). Therefore, pure RB fractions should be non-infectious whereas EB fractions should be capable of infecting cells. Purified RBs as well as EBs were diluted 100-fold and 1,000-fold in DMEM supplemented with 10% FBS. Glass tubes containing cover slips with HEp-2 cells were infected with dilutions of the RB or EB fractions and incubated under standard conditions for 24 h. After fixation in methanol,

cover slips were stained and evaluated using light or fluorescence microscopy as described in sections **4.3.3** – **4.3.5**.

4.3.10 RNA isolation

The RNA extraction from samples obtained at different stages during RB and EB separation as well as from purified forms was performed to document the purification process. 500 µl TRIsure™ (Bioline) reagent was heated to 65°C. The pellets were then resuspended in the TRIsure™ and incubated at 65°C and 550 rpm (Thermomixer® comfort) until complete pellet lysis followed by cooling down on ice. For RNA extraction, 100 µl trichlormethan/chloroform were added, the tubes were shaken vigorously for 15 s three times and centrifuged at full speed and for 3 min. After centrifugation, a phase separation occurred. RNA was solved in the upper aqueous phase which was transferred carefully into a new reaction tube. 1/10 Vol. sodium acetate (3M; pH = 5.2) as well as 3 Vol. 100% ethanol were added and the mixture was incubated at -20°C over night to precipitate RNA. Subsequent centrifugation steps were carried out at full speed and 4°C. Tubes were centrifuged for 40 min to pellet the RNA precipitate. Then, the supernatant was removed, 750 µl 75% (v/v) ethanol was added to the RNA precipitate and samples were centrifuged for 10 min. Afterwards, the supernatant was replaced with fresh 750 µl 75% (v/v) ethanol followed by incubation at RT for 15 min. Tubes were centrifuged for 10 min and the ethanol was aspirated. The washed pellets were dried under a laminar flow, resuspended in 50 µl nuclease-free water and incubated at RT for 15 min to dissolve the pellet completely. Finally, RNA concentration and quality were assessed by NanoDrop™ ND-1000 Spectrophotometer and Agilent Bioanalyzer 2100 measurements. Both measurements were performed according to the manufacturer's protocol. Using the same protocol, RNA was, also extracted from *E. coli* and HEp-2 cells, which served as prokaryotic and eukaryotic RNA control, respectively.

4.3.11 DNase I digest

For each replicate, 10 µg RNA were mixed with 5 U DNase I, 20 µl 10x DNase I reaction buffer and 2 µl SUPERase-In™. The reaction mixtures were filled up to a volume of 200 µl with nuclease-free water and incubated at 37°C for 45 min. Directly after DNase I digest, a size exclusion of RNA molecules smaller than 200 nucleotides (nts) was performed with the RNA Clean and Concentrator™-5 kit as recommended by the manufacturer. Cconcentration and quality of RNA was assessed by IMPLEN NanoPhotometer® and Agilent Bioanalyzer 2100 measurements according to the manufacturer's instructions. The successful depletion of DNA from the samples was controlled by qPCR.

4.3.12 Control of DNA removal by qPCR

DNase I treatment was validated by qPCR based on the detection of a 117 bp *16S-rRNA* fragment of *C*. *psittaci* (Binet and Maurelli, 2005), a 201 bp *ompA* fragment of *C*. *abortus* and a 150 bp *recA* gene fragment of *W*. *chondrophila* (Barkallah *et al.*, 2013). In order to determine the amount of genomic DNA (gDNA) before and after the DNase I digest the *16S-rRNA*, *ompA* and *recA* gene fragments were amplified. The qPCRs were performed in the StepOnePlus™ Real-Time PCR System as technical triplicates in 10µl reaction volume that contained 5µl 2x SsoFast™ EvaGreen® Supermix with Low ROX. The cycling profile for qPCR comprised initial denaturation at 98°C for 3 min, then 40 cycles of 3s at 95°C and 30 s at 55.5°C (*recA*); 56.0°C (*16S-rRNA*); 60.0°C (*ompA*) followed by a melt curve analysis. Each run included also no template controls (NTCs) containing sterile water instead of DNA. An important part of qPCR assays is the primer optimization (Nolan, Hands and Bustin, 2006). This revealed the optimal final primer concentrations for the *16S-rRNA* (200 nM forward and 300 nM reverse), *ompA* (500 nM forward and 300 nM reverse) and *recA* (200 nM forward and 200 nM reverse). The amount of DNA present in the RNA sample before and after DNase I digestion was determined by qPCR with serial dilutions of RNA. As a reference for normalization, 10^6 amplicon copies purified *16S-rRNA*, *ompA* and *recA* PCR products were used. The relative quantification of corresponding

gene fragment DNA between DNase I treated and non-treated RB and EB samples was performed according to the ΔΔCt method (Schmittgen and Livak, 2008).

4.3.13 Ribo-Zero™ treatment

To remove prokaryotic as well as remaining eukaryotic and mitochondrial rRNAs, the Ribo-Zero™ kit was applied according to the ScriptSeq™ Complete Gold Kit (Epidemiology) protocol. For enhanced precipitation of RNA, 1 µl GlycoBlue™ was added and samples were incubated at -20°C overnight. Precipitation, washing and desalting of RNA was carried out as described in section **4.3.10**. Successful rRNA removal and the total RNA content were determined with the Agilent 2100 Bioanalyzer system.

4.3.14 TEX treatment

Differential RNA-Seq relies on the splitting of RNA. Subsequently, one part is selectively enriched for primary transcripts using Terminator™ 5'-Phosphate-Dependent Exonuclease (TEX) and compared to a non-enriched fraction. 25 µl of RNA were mixed with 3 µl 10 x reaction buffer A, 1 µl SUPERase-In™ and 1 µl TEX (1 U/µl) or water. Then, samples were incubated at 30°C for 60 min followed by P/C/I extraction. Precipitation was enhanced by addition of 1 µl GlycoBlue™. After TEX treatment, the integrity of RNA was monitored via the Agilent 2100 Bioanalyzer system and samples were stored at -80°C.

4.3.15 cDNA library synthesis and sequencing

For the cDNA library synthesis 5 ng RNA were applied to ScriptSeq™ Complete Gold Kit (Epidemiology) according to manufacturer‘s instructions. For mutliplexing of cDNA libraries the ScriptSeq™ Index PCR Primers were used (Illumina, San Diego, USA) and final libraries were purified using the AMPure® XP magnetic beads according to the protocol of the ScriptSeq™ Complete Gold Kit. The size distribution and absence of primer dimers were evaluated using the Agilent 2100 Bioanalyzer system. If residual primer dimers could be observed with ~90 nt size, the cDNA libraries were purified repeatedly until sufficient primer removal. Final libraries were sequenced by StarSEQ GmbH (Mainz,

Germany) using a NextSeq 500 (Illumina, USA) platform and the 150 bp paired-end protocol.

4.3.16 Validation of RNA-Seq results by RT-qPCR

For each biological replicate of *C. psittaci*, *C. abortus* and *W. chondrophila* EBs and RBs, 100 ng DNAse I-digested RNA was reverse transcribed using iScript™ Reverse Transcription Supermix (Bio-Rad) according to manufacturer's protocol. For RT-qPCR, cDNA was diluted 1:5 in 10 mM Tris-HCl (pH 8.0), 0.1 mM EDTA. RT-qPCRs were performed as described in section **4.3.12** and 200 nM forward and reverse primers (**Table 6**). The chlamydial *16S rRNA* was previously applied in a study as a housekeeping gene (Belland, Zhong, *et al.*, 2003). However, it was not among the best candidates for stable expression in EBs and RBs of *C. psittaci*, *C. abortus* and *W. chondrophila*. Instead, a housekeeping index was defined with the geometric mean of the most stably expressed housekeeping candidates. These were *pmpA* and *groEL* for genes shared between *C. psittaci* and *C. abortus* and solely *hrtA* for genes shared among all three species. Gene expression was quantified using the ΔΔCt method (Schmittgen and Livak, 2008).

4.4 Bioinformatics analyses

4.4.1 Quality assessment, trimming and alignment of reads

Raw reads were obtained from StarSEQ in fastq format and evaluated using FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Trimming of low quality and adapter contaminations was performed using Trimmomatic (Bolger, Lohse and Usadel, 2014). Because of the overall short length distribution of cDNA libraries most paired end reads were overlapping and these were merged using PEAR software in default mode (Zhang *et al.*, 2014). The assembled 36–292 bp single-end reads were aligned to human (hg19), mitochondrial (NC_012920.1) and corresponding chlamydial (NC_017292, NC_004552 or NC_014225) reference genomes using Bowtie2 in "--very-

sensitive" mode (Langmead and Salzberg, 2012). For *W. chondrophila* 2032/99, only the draft genome was publicly accessible and therefore read alignment was performed to the closest relative, *W. chondrophila* WSU 86-1044 (99% sequence identity), for which the complete genome sequence was available. Alignment to the human genome was performed to determine the number of unaligned reads, which was used to assess the similarity of the sequenced strains to the public available genome sequences. Several *C. psittaci* strains harbor a conserved plasmid and therefore previously unaligned reads of *C. psittaci* 02DC15 were assembled using trinityrnaseq-2.0.4 (Grabherr *et al.*, 2011). This revealed a plasmid that is identical to *C. psittaci* 6BC (NC_017288.1). Raw count tables were generated using the GenomicAlignments v1.4.2 package (Lawrence *et al.*, 2013), whereby chlamydial rRNA and tRNAs were excluded from analyses to compensate for different efficiencies in the size selection and rRNA depletion steps.

4.4.2 Intra-species differential gene expression analysis

Differential gene expression analysis was performed using DESeq2, which normalizes sequencing depth between samples using a size factor that allows for inter-sample comparisons (Love, Huber and Anders, 2014). DESeq2 adjusts p-values for multiple testing using the FDR procedure (Benjamini *et al.*, 2001). For the intra-species comparison of differentially expressed genes (DEGs) upon TEX treatment and in EBs and RBs, a p-value cutoff <0.05 adjusted for multiple testing (p_{adj}) was applied.

4.4.3 Inter-species DEGs

For the identification of species-specific and homologous genes, an all-vs.-all comparison of *C. psittaci*, *C. abortus* and *W. chondrophila* genomes was performed using RAST (Aziz *et al.*, 2008). Identified bidirectional hits were filtered for hits ≥30% total DNA sequence identity. In case of multiple matches only the best was retained. The best reciprocal hits by these criteria were considered as homologs. For the inter-species comparison of DEGs DESeq2 was used and a significance threshold p_{adj} <0.01 and an absolute log2-fold change >1.0 were applied.

4.4.4 Assignment of genes to functional categories

Chlamydial genes were assigned to crude categories reflecting fundamental divisions of biological function using the RAST database (Aziz *et al.*, 2008). The assigned RAST categories were “Amino Acids and Derivatives”; “Carbohydrates”; “Cell Division and Cell Cycle”; “Cell Wall and Capsule”; “Cofactors, Vitamins and Prosthetic Groups”; “DNA Metabolism”; “Fatty Acids, Lipids and Isoprenoids”; “Membrane Transport”; “Miscellaneous”; “Nitrogen Metabolism”; “Nucleosides and Nucleotides”; “Phosphorus Metabolism”; “Potassium Metabolism”; “Protein Metabolism”; “Respiration”; “RNA Metabolism”; “Stress Response”; “Sulfur Metabolism”; and “Virulence, Disease and Defense”. To extend the list of categories, predictions for Incs, Omps, Pmps, plasticity zone (PZ) and T3A genes were assembled from the literature (Thomson et al., 2005; Bertelli et al., 2010; Voigt, Schöfl and Saluz, 2012). This yielded the additional categories “Inclusion Membrane Proteins”; “Outer Membrane Proteins”; “Polymorphic Membrane Proteins”; “Genes within Plasticity Zone”; and “T3S Apparatus”. In addition to the RAST categories and the literature search, T3S effectors were included. In *C*. *psittaci* and *C*. *abortus*, potential T3SE were predicted using EffectiveT3 (Arnold *et al.*, 2009), with a cutoff >0.9999, and compared to a machine learning approach (Voigt, Schöfl and Saluz, 2012). Genes predicted to be type III secreted by both methods were considered T3SE. For *W*. *chondrophila*, only EffectiveT3 predictions were available.

DEGs within each subcategory were tested for over- and under-representation using the two-sided Fisher’s exact test, G-test and hypergeometric test and p-values were FDR corrected (Benjamini *et al.*, 2001). All tests are used to determine whether there is a significant difference between the expected and observed frequencies. The G-test was included because its more robust for small samples than the Chi-squared test. All statistical tests were conducted with the corresponding functions implemented in R (version 3.4.4; RDC Team, 2013). Categories with fewer than two genes were omitted.

4.4.5 Identification of TSSs in *Chlamydia*

In order to obtain reliable TSS annotations for *Chlamydia*, three different automated prediction methods were applied, i.e. TSSpredator-1.06 (Dugar *et al.*, 2013), TSSAR-1.0.1 (Amman *et al.*, 2014) and TSSer-1.0 (Jorjani and Zavolan, 2014). All tools need aligned dRNA-Seq data as input. In order to evaluate the TSS annotation performance of the tools individually and in combination a published dRNA-Seq data set of *E. coli* (Thomason *et al.*, 2015) was reanalyzed. For optimization the first 201,484 nts (~5%) of the *E. coli* genome were manually curated based upon the enrichment of reads towards the TSS in the TEX treated libraries (**Figure 2**). These manual determined TSSs were used for performance evaluation of the tools, whereby a TSSs had to be detected in at least two replicates of a species to be accounted as a valid hit. TSSpredator parameters (*height*, *height reduction*, *factor*, *factor reduction*, *enrichment factor*, *processing site factor* and *base height*) were optimized employing a genetic algorithm approach, that was trained on the manual curated set of predictions and implemented in the ANNOgesic pipeline (Yu, Vogel and Förstner, 2018). In TSSAR only two parameters (*minPeak* and *window size*) for annotation had to be optimized. Thereby *minPeak* is the enrichment of reads forming a pronounced peak in the TEX treated compared to the non-treated library. This parameter was successively increased from 3 to 24 (by 1). The *window size* for detecting these peaks was increased from 500 to 4,000 in 500, equally sized, steps. TSSer provides 1,200 result tables with varying parameters, i.e. *filtered enrichment*, *local, norm exp* and *clustered linkage*. In order to select the best results, parameters with same or better sensitivity (Sn) as the default setting and higher specificity (Sp) were selected.

In *C. psittaci, C. abortus and W. chondrophila* the same optimization for each TSS annotation tool was carried out. However, only the first 50,000 nts of the genomes were manually inspected for TSSs, because of the shorter genome length. In addition, the *C. pneumoniae* dRNA-Seq data set (Albrecht *et al.*, 2011) was reanalyzed because an elaborated prediction of TSSs in intergenic and antisense regions has not been performed in the original study. Due to the lower sequencing depth of the

C. pneumoniae libraries (**Supplementary Table 1**), the first 140,000 nts had to be inspected in order to find enough TSSs for performance evaluation. The published dRNA-Seq data of *C. trachomatis* (Albrecht *et al.*, 2010) did not yield significant results upon reanalysis, because of the low sequencing depth (<10,000 reads mapping to non-rRNAs locations per library) and there was only one replicate of EBs and RBs available. In total 56 dRNA-Seq libraries were analyzed.

4.4.6 Identification of conserved motifs in chlamydial promoters

In order to identify consensus structures, promoter sequences -1 to -40 nts upstream of all 3,561 identified chlamydial TSSs were extracted and searched for conserved motifs using FIMO (Grant, Bailey and Noble, 2011). As input, the position weight matrices (PWMs) derived from *C. pneumoniae* σ^{66} and another prominent motif (CTTG-N_{20}-TAT) were generated (Albrecht *et al.*, 2011). For the other two known chlamydial σ factors (σ^{54} and σ^{28}) PWMs of *E. coli* were downloaded from prodoric2 (Eckweiler *et al.*, 2018) and searched using FIMO. Sequence motifs were generated using WebLogo (Crooks *et al.*, 2004). In addition to the motif scanning approach, a *de novo* motif discovery analysis using MEME (Bailey *et al.*, 2015) was performed.

4.4.7 ORF, promoter and 5'UTR sequence comparison

Knowing the TSS positions it was possible to extract beside the promoter also the 5'UTR sequences. A subsequent comparison of ORF, promoter and 5'UTR sequences was performed using the R packages seqinr 3.4-5 version (Charif and Lobry, 2007) and Biostrings version 2.46.0 (Pages *et al.*, 2019).

5 Results

5.1 Purification of EBs and RBs

Isolation of integer EBs and RBs was an important prerequisite to study their transcriptomes. To examine the purity of the fractions, TEM imaging, RNA integrity measurements and reinfection assays were performed. Exemplary results of purified EBs and RBs are shown in **Figure 3**. For comparison, the crude extract taken from a *C. abortus* sample before it was layered on top of the iodixanol-gradient was examined too. The crude extract showed abundant granular and vesicular host cell residuals and only occasionally *Chlamydia* could be observed (**Figure 3A**). After purification, *C. abortus* (**Figure 3B** and **C**), *C. psittaci* (**Figure 3D** and **E**) and *W. chondrophila* (**Figure 3F** and **G**) were enriched and RB fractions were virtually free of EBs and *vice versa*.

Crucial for RNA-Seq experiments is the isolation of high quality RNA. Therefore, integrity of isolated RNA was controlled and compared to the input (*Chlamydia* infected HEp-2 cells) and crude extract (**Figure 4**). Both, input and crude extract showed a characteristic pattern of two dominant bands corresponding to the eukaryotic 28S and 18S rRNA and two minor bands, representing the chlamydial 23S and 16S rRNAs, respectively. In contrast, the purified RB and EB fractions contained exclusively the chlamydial 23S and 16S rRNA bands (**Figure 4**).

The third indicator for sufficiently pure EB and RB fractions was obtained by reinfection assays (**Table 7**). EBs are the only infectious chlamydial state and indeed, IFU/ml derived from isolated EBs of all three pathogens were much higher compared to RB fractions (>100:1). The ten times lower efficiency in *W. chondrophila* IFU/ml yield was due to the ten times lower MOI of 0.5.

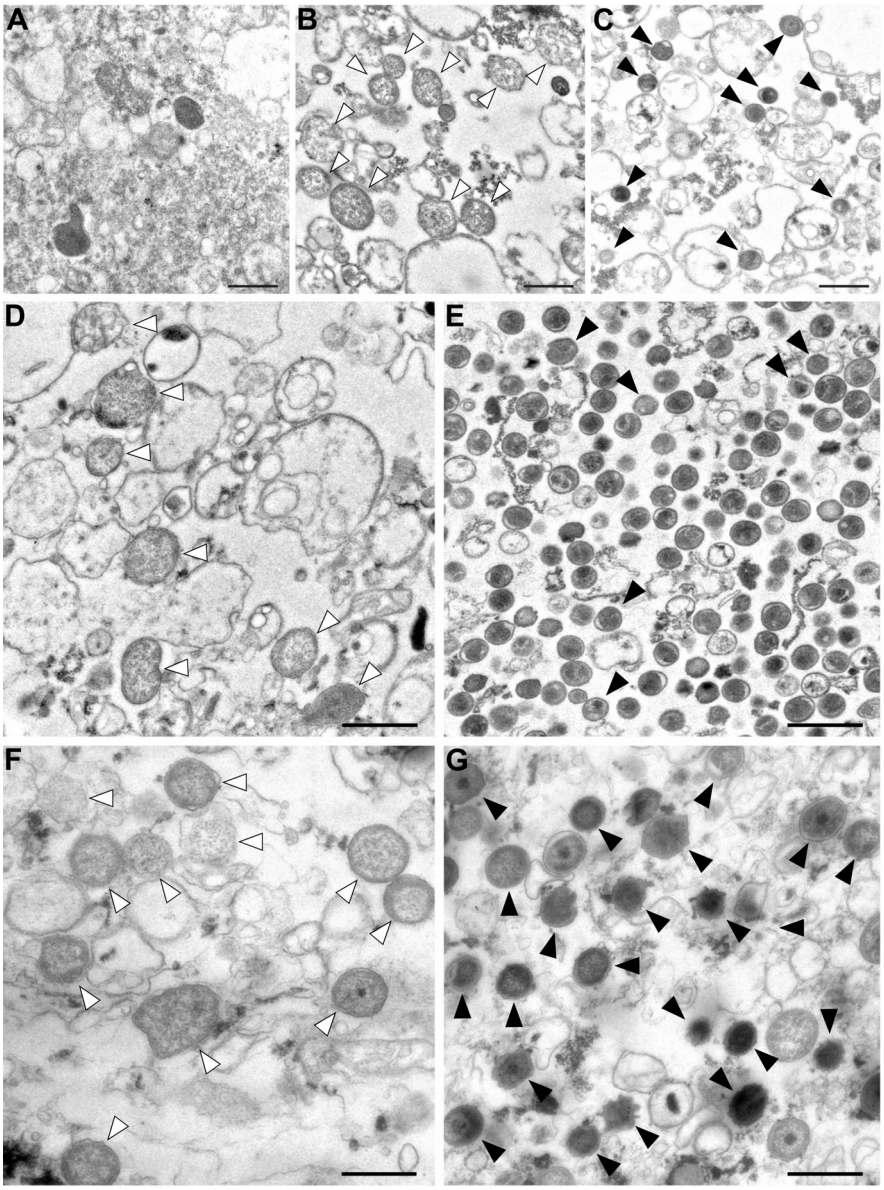


Figure 3. TEM images of purified RBs and EBs. HEp-2 cells were infected with *C. abortus* (A, B and C), *C. psittaci* (D and E) and *W. chondrophila* (F and G) followed by purification of RBs and EBs. Representative TEM images of crude extract (A), purified RBs (B, D and F) and EBs (C, E and G) are shown. In the crude extract, some *Chlamydia* beside abundant granular and vesicular host cell debris can be seen. Host cell contaminants are depleted and *Chlamydia* are concentrated upon purification. RBs appear as round to irregular forms with evenly dispersed granular material and a diameter of 0.5 - 1.0 µm (white arrowheads). EBs are smaller, round and contain powdery electron dense cell material (black arrowheads). The scale bar of each image represents 1 µm.

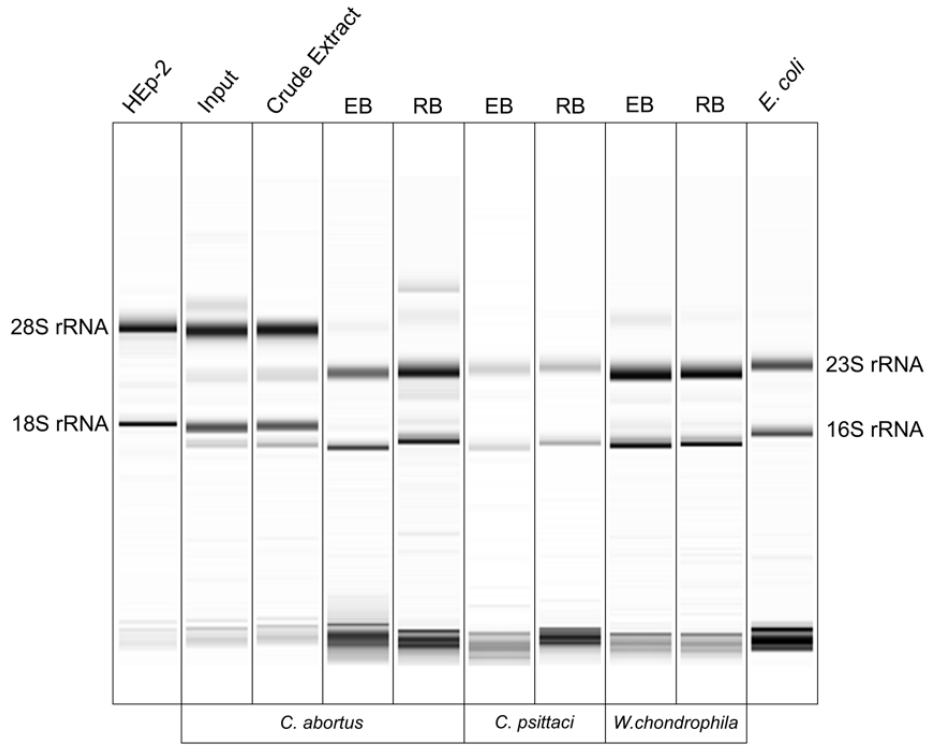


Figure 4. Host RNA is depleted and chlamydial RNA is enriched upon RB and EB purification. Gel-like images of isolated total RNA created by Agilent Bioanalyzer software are shown. Prominent eukaryotic and prokaryotic band patterns are represented by HEp-2 and *E. coli* RNA, respectively. Input and crude extract exhibit dominant signals for eukaryotic 28S as well as 18S rRNAs and only faint signals for chlamydial 23S and 16S rRNAs. In purified EB and RB fractions of *C. abortus*, *C. psittaci* and *W. chondrophila* eukaryotic rRNA band patterns are absent, representing the depletion of host cell material and enrichment of *Chlamydia*.

Table 7. Mean value of infectious *Chlamydia* in RB and EB fractions.

Spcecies	Type	IFU ml^{-1} ± SD	EB : RB
C. psittaci	RB	1.6 x 10^6 ± 6.1 x 10^5	146:1
	EB	2.4 x 10^8 ± 6.9 x 10^7 **	
C. abortus	RB	1.4 x 10^6 ± 1.1 x 10^6	206:1
	EB	2.9 x 10^8 ± 2.9 x 10^8 *	
W. chondrophila	RB	2.2x 10^5 ± 0.4 x 10^5	124:1
	EB	2.7 x 10^7 ± 8.8 x 10^6 **	

Number of inclusion forming units (IFU) per ml from purified RBs and EBs is represented as mean ± SD of independent replicates (*C. psittaci* n = 4; *C. abortus* and *W. chondrophila* n = 3). Statistics were conducted using Student's one-tailed t-test for independent samples.

* represents p-value < 0.05

** represents p-value < 0.01

5.2 RNA processing for sequencing

Depletion of rRNAs and subsequent enrichment of primary transcripts by TEX treatment were not offered by any company at the time the experiments were carried out. Therefore, RNA processing for sequencing, which included DNase I digest, size selection of transcripts >200 nts, rRNA depletion, TEX treatment and cDNA library synthesis had to be established. The integrity of RNA was monitored by Agilent 2100 Bioanalyzer measurements after each processing step (**Figure 5**). DNase I digested RNA showed a very specific distribution of transcripts. Two peaks for the 16S and 23S rRNAs could be detected with sizes of about 1,600 nts and 3,000 nts, respectively. In addition, a wider accumulation of transcripts from 50 to 400 nts containing the 5S rRNA and tRNAs was observed (**Figure 5**). Removal of DNA by DNase I treatment was controlled by qPCR and quantified by the ΔΔCt method. The average removal of DNA was 236 fold, demonstrating a depletion of DNA by orders of magnitude. DNase I digested RNA was next applied to the Ribo-Zero™ for rRNA depletion and the characteristic 16S and 23S rRNA peaks were absent after the treatment (**Figure 5**). At net, the rRNA depleted samples were split and one part was enriched for primary transcripts by selective digestion of transcripts carrying 5'P ends. The other part was incubated in reaction buffer alone. Finally, cDNA libraries were generated and after PCR amplification sizes ranged from 150 to 600 bp with a maximum at around 250 nts (**Figure 5**). In total 40 cDNA libraries were prepared for sequencing using this protocol.

5.3 Quality control, trimming and alignment of reads

Sequencing raw data was obtained as fastq files, which were next checked for quantity and quality of the reads. Most samples had ~5,000,000 million reads with a minimum of 1,992,001 in the *W. chondrophila* EB_2_+TEX and maximum of 10,344,925 reads in the *C. abortus* EB_3_-TEX library (**Figure 7**). Generally, sequencing quality was high in all samples and therefore, in **Figure 6** only the quality assessment of 12 *C. abortus* EB libraries (both paired mates) before and after trimming is shown. The quality of data

dropped towards the read end (**Figure 6A**), which is common for data derived from Illumina sequencing platforms. Thereby, a Phred Score of 40 represents 99.99% base call accuracy 30 equals 99.9% and so on. In addition to the decreasing accuracy, the raw data showed increasing adapter contents towards the end of the reads (**Figure 6B**). This is a result of read through of ligated sequencing adapters during sequencing.

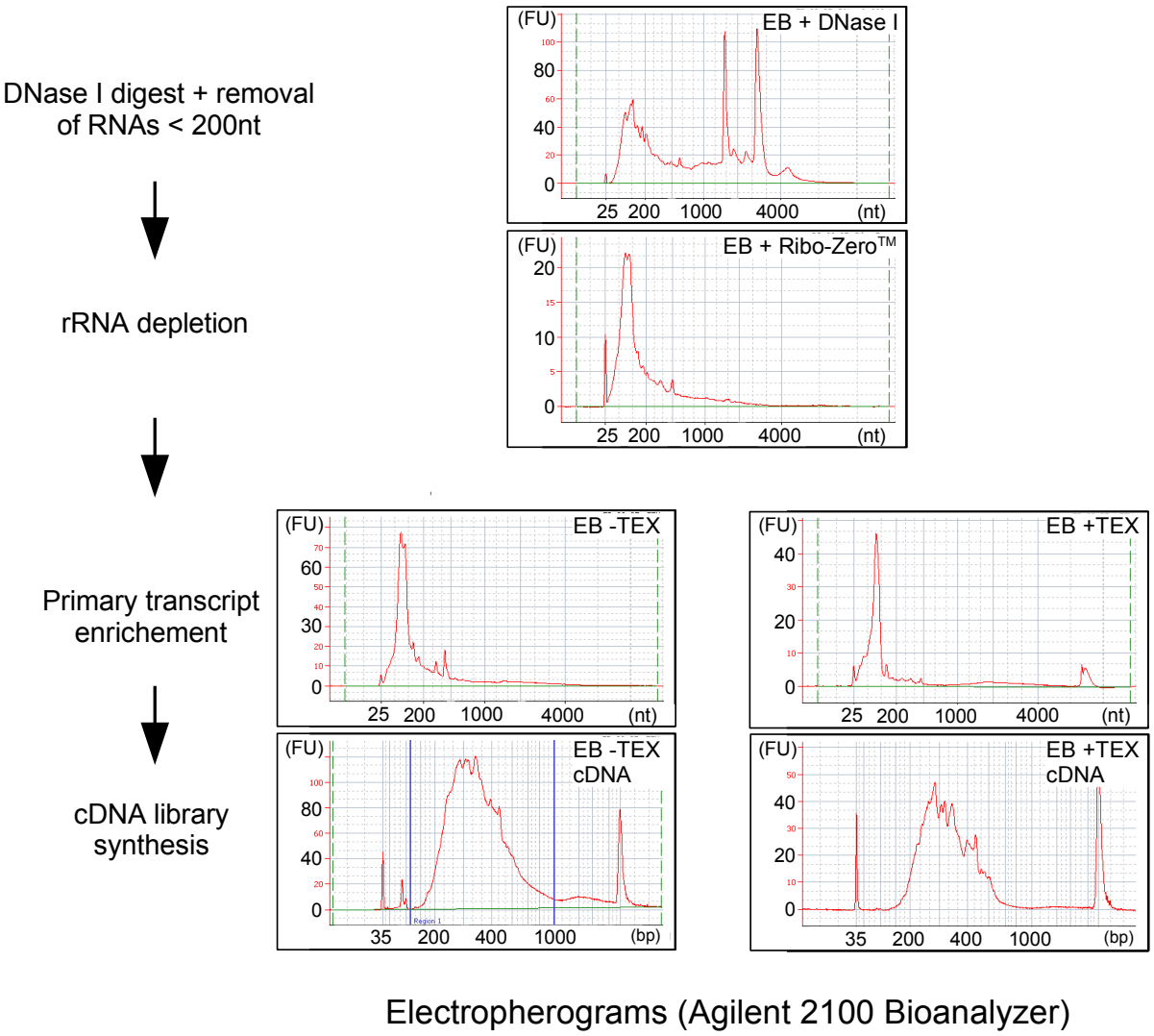


Figure 5. RNA processing and synthesis of cDNA libraries. The workflow comprised DNase I digestion, rRNA depletion with Ribo-Zero™, enrichment of primary transcripts by TEX treatment and cDNA library preparation. After each processing step the integrity and size distribution of nucleic acids was monitored by Agilent 2100 Bioanalyzer measurements.

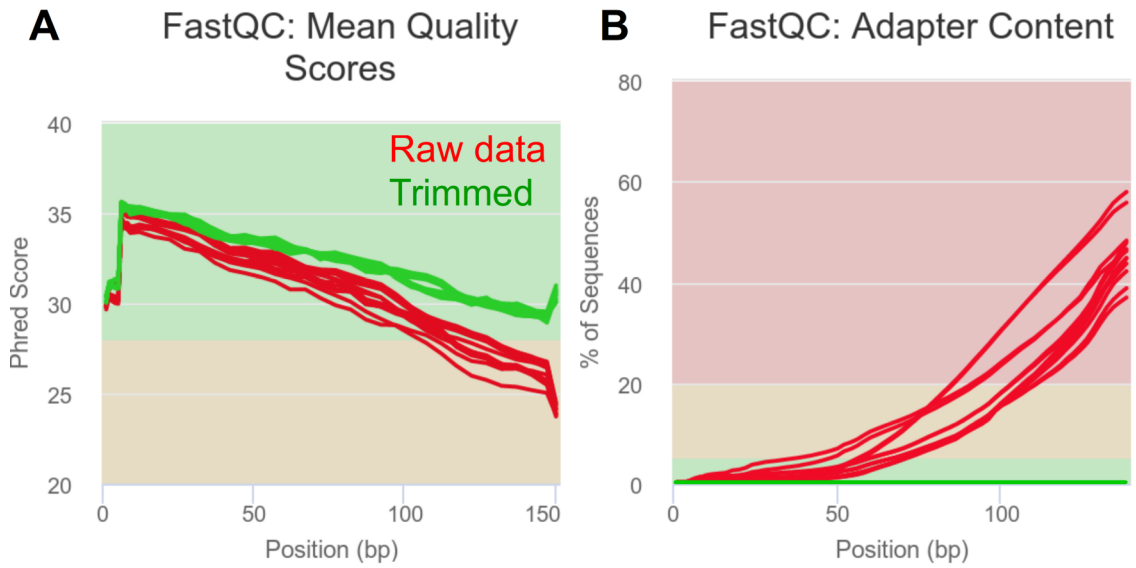


Figure 6. Trimming low quality and technical sequences of *C. abortus* EB libraries. Quality (A) and adapter contents (B) of libraries were assessed by FastQC before (red) and after (green) trimming. The combined plots were generated using MultiQC (Ewels *et al.*, 2016).

After trimming of low quality and adapter sequences only high quality reads remained and the adapters were completely clipped. Trimming resulted in 6% – 21% discarded reads whereby in the *W. chondrophila* samples more reads had to be removed (**Figure 7**; p-value <0.001). Due of the large fraction of adapter sequences in the reads (**Figure 6B**) and the short size distribution of cDNA libraries (**Figure 5**) a large amount of overlapping paired end reads was expected. Because of that, paired end reads were merged, which revealed that indeed more than 90% were overlapping (**Figure 7**). The merging resulted in longer single end reads, and these were aligned to the human, mitochondrial and corresponding chlamydial reference genomes. This was done to estimate, first the fraction of reads aligning to the chlamydial genome, second host derived contaminations, and third to assess the fraction of unaligned reads between the *C. psittaci* genome, which was sequenced on behalf of the CMB department (Schöfl *et al.*, 2011) and the public available *C. abortus* and *W. chondrophila* genomes. On average, 2,700,000 reads mapped to the chlamydial genome, which ranged from

753,445 in the *W. chondrophila* RB_1_-TEX to 4,760,298 in *C. abortus* EB_2_-TEX library (**Supplementary Table 1**).

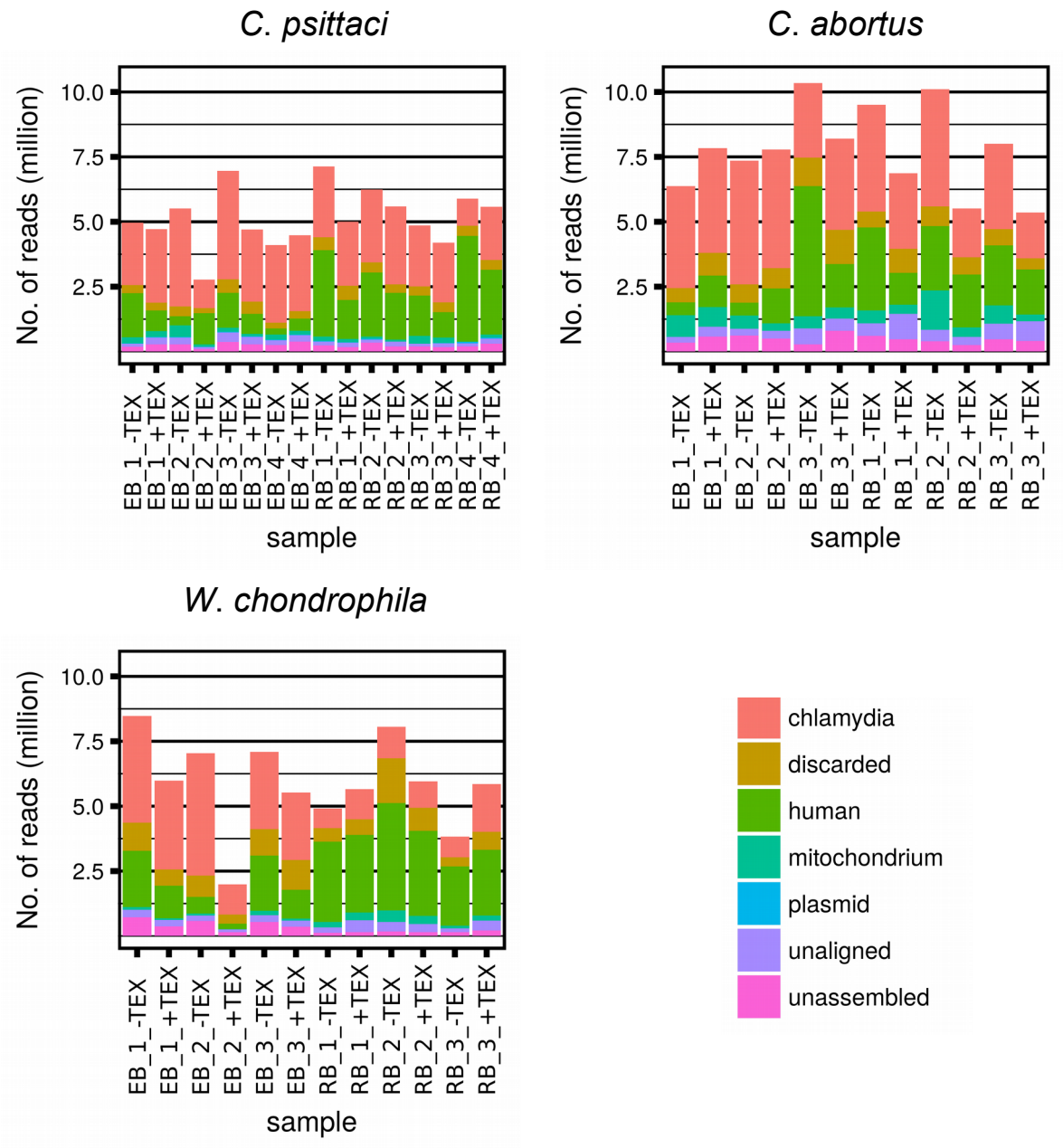


Figure 7. Data trimming, assembling and alignment results. Sequencing reads of each *C. psittaci*, *C. abortus* and *W. chondrophila* library were quality filtered, assembled and aligned. The fraction of "discarded", represents reads removed upon trimming and "unassembled", reads that could not be merged. The proportion of reads that neither mapped to host nor chlamydial genomes are depicted as "unaligned".

Generally, in the RB samples more reads mapped to the human genome in all species (p-value <0.001). Consequently, this indicates more host derived contaminations were present at the 30%/37% Visipaque® inter-phase, probably caused by pieces of rough endoplasmatic reticulum (**Figure 3**). Only a small fraction of reads did neither map to host nor chlamydial genomes i.e. 3.7% in *C. psittaci*, 4.6% in *C. abortus* and 6.5% in *W. chondrophila* samples, respectively (**Figure 7**). Considering that bowtie2 fails to align reads spanning eukaryotic exon-exon boundaries, the fraction of reads that not align is even smaller. In addition to the *C. psittaci* 02DC15 genome, a ~7.5 kb plasmid was found by *de novo* assembly of unaligned reads, which is identical to plasmids of other strains of that species. For completeness alignment results to the plasmid are shown in **Figure 7** even though only 0.85% of reads mapped to it.

5.4 Read distribution on genomic locations and coverages

Generally, numbers of mRNA copies derived from individual genes can differ by several orders of magnitude. Therefore, detecting low abundant transcripts relies on a sufficient sequencing depth. The number of reads mapping to mRNAs can be increased by selective removal of rRNAs (which represent 80-95% of bacterial transcriptomes) prior to cDNA library construction (Giannoukos *et al.*, 2012). For this purpose Ribo-Zero™ was applied, which removes bacterial as well as remaining eukaryotic and mitochondrial rRNAs. Ribosomal RNA removal resulted in a mean genome coverage of 301 (**Supplementary Table 1**), but more important for this study was the fraction of reads aligning to CDSs. For these a mean coverage of 155 were achieved that ranged from 34 in *W. chondrophila* RB_1_-TEX up to 448 in *C. abortus* RB_1_-TEX (**Supplementary Table 1**). For detection of all but a few of the lowest expressed genes in diverse bacteria growing under a variety of conditions 5 – 10 million non-rRNA reads have been suggested to be sufficient (Haas *et al.*, 2012). In average ~1.9 million reads aligned to CDSs and considering the small size of chlamydial genomes (~1 Mb) the sequencing depth was

very adequate. This is also represented by the number of genes with zero counts that were only 3/975 in *C. psittaci*, 2/933 in *C. abortus* and 52/1840 in *W. chondrophila*.

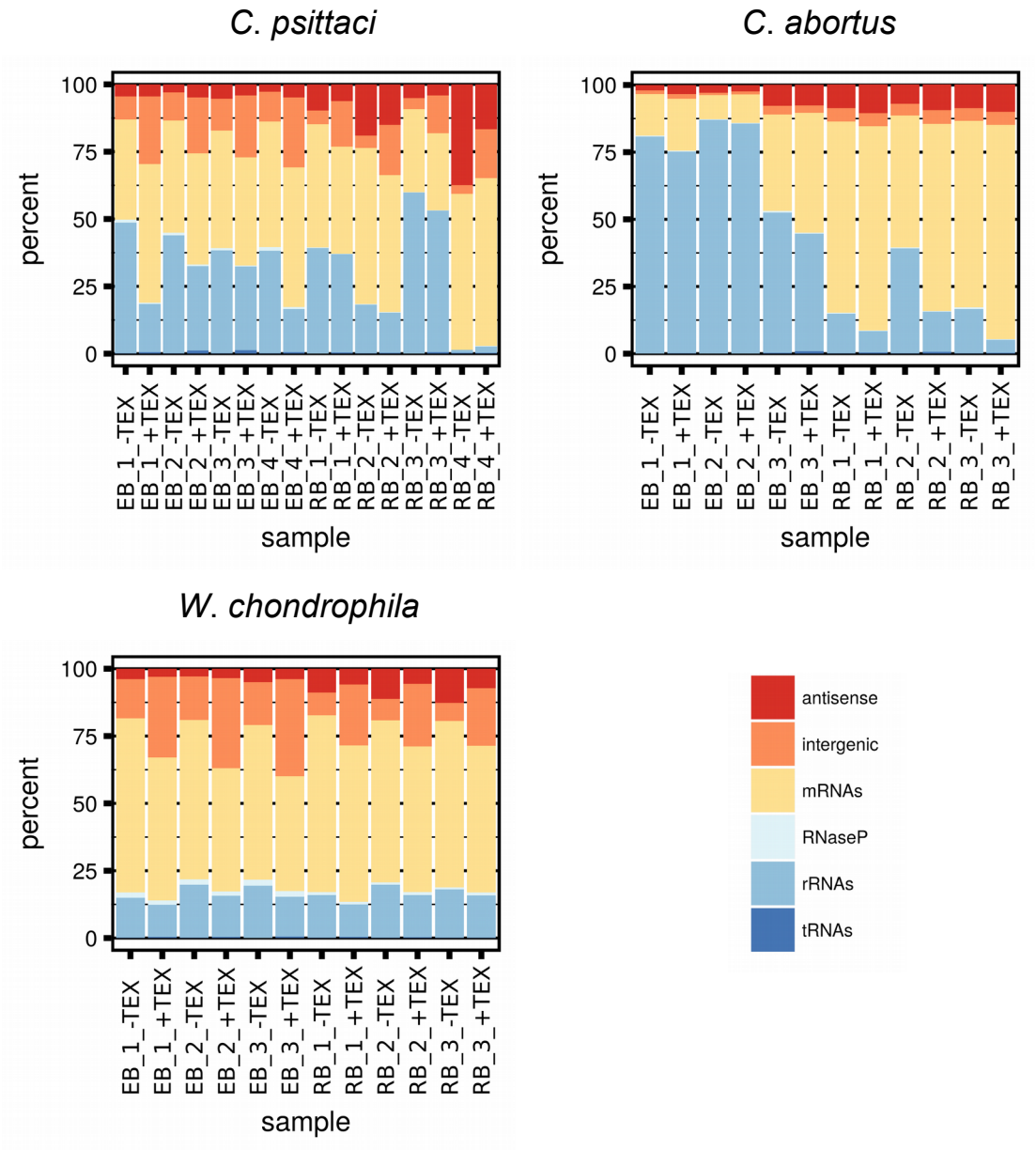


Figure 8. The distribution of reads mapping to different genomic locations. Most reads mapped to mRNAs after depletion of rRNAs by Ribo-Zero™. Transcripts located in intergenic regions can either be sRNA candidates, part of UTRs or unannotated coding genes.

After rRNA removal, most reads mapped to mRNAs (**Figure 8**). In addition, the fractions of reads mapping to other genomic locations like intergenic and antisense were

assessed. The transcripts in intergenic regions can be either sRNA candidates, part of UTRs or unannotated coding genes. Antisense transcripts might be involved in post transcriptional regulation of gene expression. Ribo-Zero™ treatment resulted in only 30% of reads on average aligning to rRNAs but the fraction varied between samples (**Figure 8**). The amount of rRNA was further decreased by TEX treatment (p-value <0.001), which was expected because rRNAs carry 5'P ends making them vulnerable for TEX digestion. Generally, only few reads mapped to tRNAs (0.26% – 0.41%) probably due to the size exclusion of transcripts <200 nts. Antisense and intergenic transcripts were detected in a range from 3.3% - 40.7% depending on the ratio of initially depleted rRNAs. Interestingly, the proportion of reads mapping antisense to known features was similar in -TEX and +TEX samples, whereas more reads mapped to intergenic regions in the +TEX samples (p-value <0.001). This fits well to the mode of action of TEX treatment, since it enriches transcripts towards the 5'-end (**Figure 2**). These are located downstream of the start codon and therefore considered as intergenic regions.

In summary, rRNA removal highly increased the number of reads mapping to mRNAs, antisense RNAs and to intergenic regions. Consequently, high coverages for all organisms were achieved (**Supplementary Table 1**).

5.5 Intra-species comparison of gene expression

5.5.1 Differential gene expression and gene set enrichment analysis

Differential gene expression analysis was performed using DESeq2. Initially the incorporation of the *C*. *trachomatis* and *C*. *pneumoniae* dRNA-Seq data was planned (Albrecht *et al.*, 2010, 2011), but reanalysis yielded non significant results due to low sequencing depth (**Supplementary Table 1**) and small number of replicates. Therefore, the differential gene expression analysis focused on *C*. *psittaci*, *C*. *abortus* and *W*. *chondrophila*. TEX digestion enriches primary transcripts harboring a 5'PPP end (**Figure 2**), which are represented by DEGs up-regulated upon TEX treatment (**Table 8**). In contrast, depleted mature transcripts harbor a 5'P end making them accessible for

mRNA decay (Celesnik, Deana and Belasco, 2007). In EBs, TEX treatment resulted in 161 (*C. abortus*) up to 536 DEGs (*W. chondrophila*). In RBs, the results were more different among the species ranging from 397 DEGs in *C. abortus* to only 61 in *C. psittaci* (**Table 8**). The reason for the smaller number of DEGs in *C. psittaci* RBs is explained by the relatively low variation between -TEX and +TEX samples. This results in combined clustering of RB samples shown in **Figure 9**. *C. psittaci* EB samples clustered more distinctly (despite EB1-TEX and EB2+TEX forming a sub-cluster). This indicates more different transcriptomes in -TEX and +TEX EB samples resulting in more DEGs (**Table 8**).

Table 8. Summary of differential gene expression analysis

Spcecies	Comparision	DEGs
C. psittaci	EB -TEX vs. EB +TEX	234 (161↑; 73↓)
	RB -TEX vs. RB +TEX	61 (60↑; 1↓)
	EB -TEX vs. RB -TEX	617 (323↑; 294↓)
C. abortus	EB -TEX vs. EB +TEX	161 (104↑; 57↓)
	RB -TEX vs. RB +TEX	397 (203↑; 194↓)
	EB -TEX vs. RB -TEX	320 (153↑; 167↓)
W. chondrophila	EB -TEX vs. EB +TEX	536 (287↑; 249↓)
	RB -TEX vs. RB +TEX	391 (238↑; 153↓)
	EB -TEX vs. RB -TEX	1029 (515↑; 514↓)

Number of differential expressed genes are shown as detected by DESeq2 with Benjamini and Hochberg adjusted p-values <0.05.
↑up-regulated genes; ↓ down-regulated genes

However, in all three species EB and RB transcriptomes were very different which is indicated by the clear clustering in **Figure 9** and the high number of DEGs, which sum up to 63%, 34% and 56% of all genes in *C. psittaci*, *C. abortus* and *W. chondrophila*, respectively (**Table 8**). In *C. abortus* RB also the -TEX and +TEX samples cluster distinctly apart (**Figure 9**), which is also reflected by the high number of 397 DEGs. Within EBs of *C. abortus* clustering is less rigorous, reflected again by the smaller number of 161 DEGs (**Table 8**). The most distinct clustering of -TEX and +TEX samples in EBs as well as RBs was found in *W. chondrophila*. Taken together, TEX treatment resulted in numerous DEGs and transcripts enriched display primary transcripts of EBs and RBs.

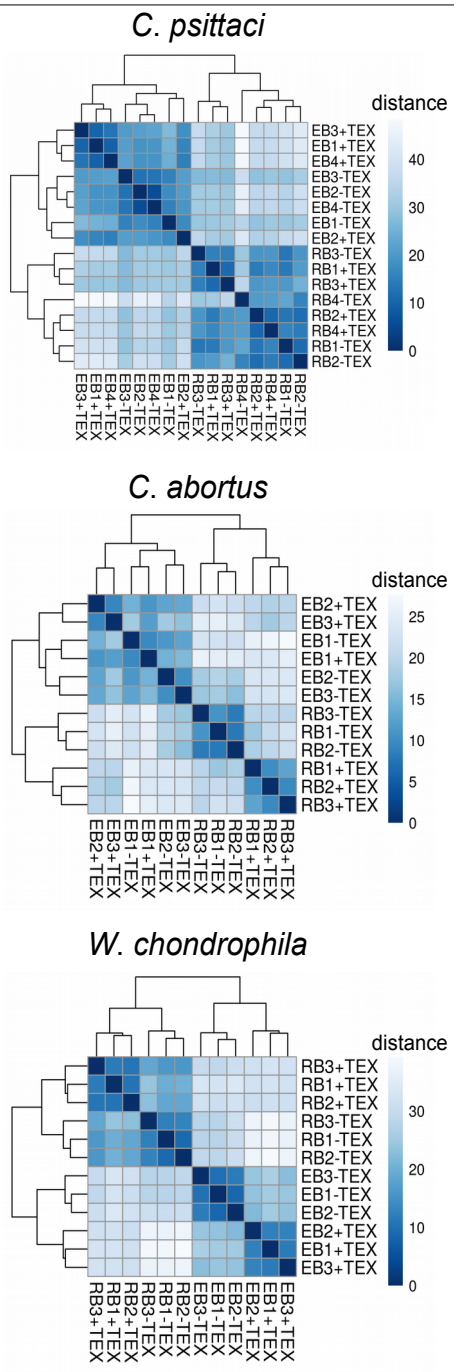


Figure 9. Heat map and clustering. The rlog transformed counts for all samples were used to calculate the euclidean distance visualized as a heatmap. Hierarchical clustering was performed using the complete-linkage method generating the dendrograms.

Furthermore, the transcriptomes of EBs and RBs are very different and one-third to one-half of all genes are differentially expressed in the developmental states (**Table 8**).

The functional categories of DEGs were tested for over- and under-representation, which revealed that categories like “Amino Acids and Derivates”, “Cofactors, Vitamins, Prosthetic, Groups, Pigments” or “RNA Metabolism” are overrepresented RBs, whereas genes involved in “DNA Metabolism” are underrepresented in EBs (**Supplementary Table 2**). This is in accordance with the biological functions of the replicative RBs and infectious EBs. Interestingly, categories enriched in *W. chondrophila* EBs are “Carbohydrates” and “Respiration”, which indicates respiratory activity in the infectious *W. chondrophila* state. Generally, TEX treatment resulted in a uniform enrichment of transcripts in the different categories, which is illustrated by the low number of over- or under-represented categories (**Supplementary Table 3**). Only genes involved in “Protein Metabolism” were over-proportional depleted upon TEX treatment in EBs of *C. psittaci* and *C. abortus*. This means that transcripts involved in “Protein Metabolism” are in the process of mRNA decay in EBs, which again fits well to the dormant state in which proteins are

preloaded and synthesis stops. Some categories like “T3S apparatus” and “Protein Metabolism” were differentially over- or under-represented in EBs and RBs of *C. psittaci*, *C. abortus* and *W. chondrophila* (**Supplementary Table 2**). This indicates that among species certain biological processes might be differentially emphasized.

5.5.2 Differential expression of virulence factors in EBs and RBs

VFs in this work are defined as proteins that aid infection, proliferation, resistance or transmission of pathogens. In *Chlamydia*, several VFs have been investigated and 52 of these (adapted from Collingro et al., 2011), which compromise virulence, immunogenic and T3SE proteins were tested for differential expression in EBs and RBs. Of these genes 30 are shared among *C. psittaci* and *C. abortus* and 22 also with *W. chondrophila*. Gene name, function, homologs and expression of the VFs are shown in **Supplementary Table 3**. Generally expression of the VFs is positively correlated between *C. psittaci* and *C. abortus* (0.407; p-value = 0.002715) but the expression was more diverse than initially expected. Up-regulated in EBs of both species are for example the chaperones *dnaK*, *groEL* and *groES* which are responsible for preventing misfolding and aggregation of protein molecules. Another class of chaperones i.e. the type III secretion chaperones have been shown to be enriched in EBs (Saka *et al.*, 2011) and were also found to be transcriptional up-regulated in the infectious state.

To confirm the RNA-Seq results, the expression of five genes shared among *C. psittaci* and *C. abortus* (*cap1*, *copB_1*, *incA*, *ompA* and *sinC*) and four shared among all three species (*dnaK*, *hctA*, *mip* and *omcB*) was validated using RT-qPCR (**Figure 10**). These homologs represented low, high as well as differentially and similarly expressed genes. The log2-fold changes of gene expression that were determined by RNA-Seq strongly correlated with the RT-qPCR results (**Figure 10**). For the five transcripts shared among *C. psittaci* and *C. abortus* (**Figure 10A**) a correlation coefficient of 0.941 (p-value = 4.895e-05) and for the transcripts shared by all three pathogens (**Figure 10B**) a correlation coefficient of 0.893 (p-value = 9.261e-05) was found.

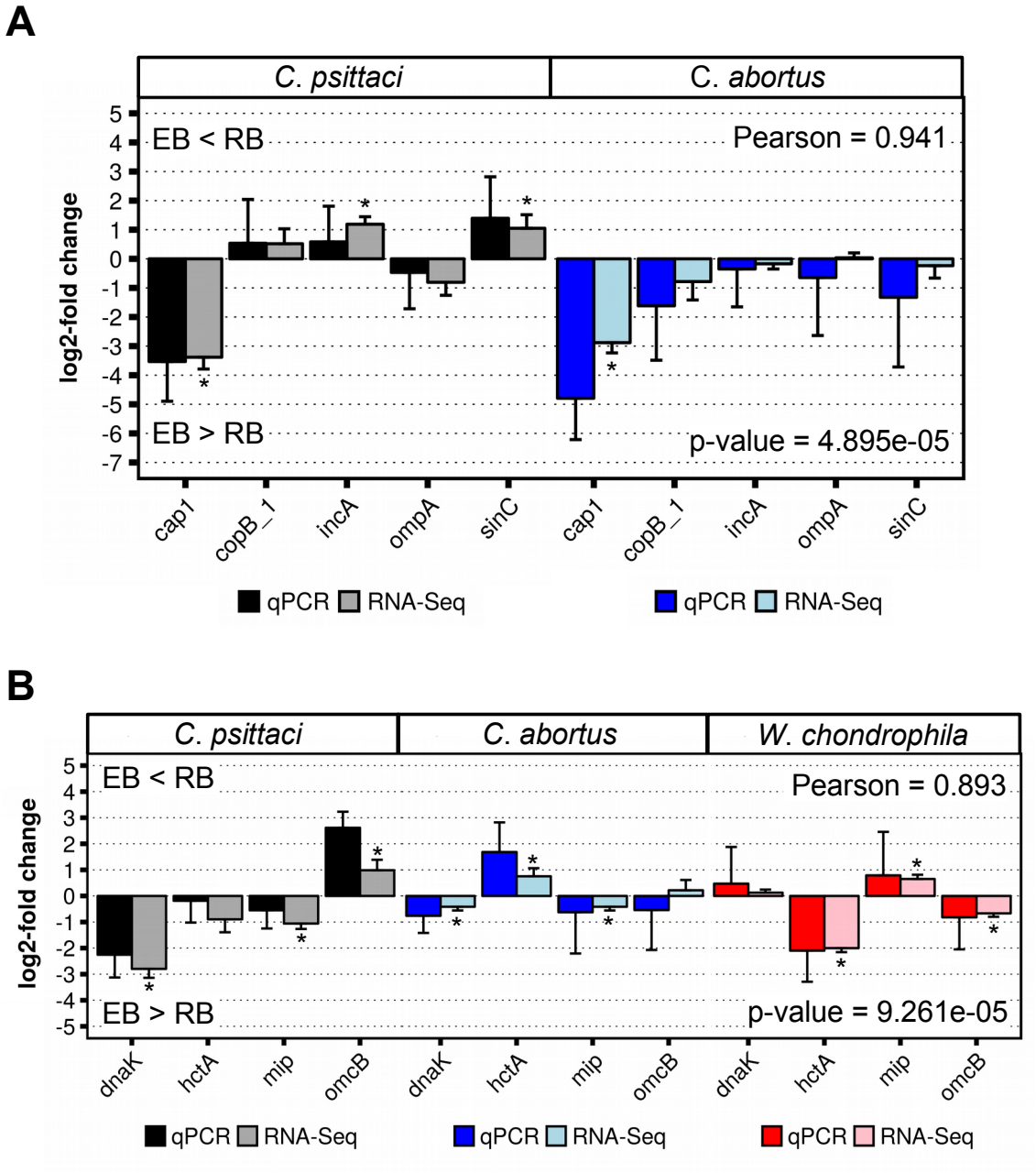


Figure 10. Correlation of the RNA-Seq and RT-qPCR results. Expression of five virulence factors shared between *C. psittaci* and *C. abortus* (**A**) and four also shared with *W. chondrophila* (**B**) was validated by RT-qPCR. The Pearson correlation coefficients of 0.941 and 0.893 demonstrated a high correlation of the two techniques. Standard deviation of log2-fold changes are indicated. Significance threshold is depicted by asterisk (*) and represents Benjamini-Hochberg adjusted p-values <0.05 of the RNA-Seq results.

5.6 Annotation of Transcription start sites

5.6.1 Optimization of TSS annotation

To note, dRNA-Seq relies on TEX treatment of RNA which leads to an enrichment of primary transcripts with a 5'PPP group (Celesnik, Deana and Belasco, 2007). Consequently, TEX treated samples are enriched for reads towards the nuclease-protected 5'PPP-end and thus a comparison to the untreated libraries allows TSS annotation at single nucleotide resolution (**Figure 2**). This basic principle of dRNA-Seq can be used for visual inspection of the data (Albrecht *et al.*, 2010, 2011; Sharma *et al.*, 2010; Kröger *et al.*, 2012) and by tools that automatically annotate TSSs (Dugar *et al.*, 2013; Amman *et al.*, 2014; Jorjani and Zavolan, 2014). The main advantage of the automated annotation is that it is much faster and can be integrated in analysis pipelines. However, computational annotation resulted in numbers of TSSs many times excessing known genes (Thomason *et al.*, 2015). Therefore, one focus of this work concerned the reliable computational annotation of TSSs and consequently, the performance of different tools had to be evaluated. For reference the first ~5% of the genomes were manually annotated for TSSs based on the enrichment of the reads towards the 5'-end (**Figure 2**).

Optimizing the TSS predictions of each tool individually and combination of tools should principally result in more reliable annotations. To test the approach, the published dRNA-Seq data set of *E. coli* (Thomason *et al.*, 2015) was reanalyzed. *E. coli* was used because another high throughput data set using a technique called Cappable-seq (Ettwiller *et al.*, 2016) for TSS annotation was available for validation. The 181 manually determined TSSs on the first ~5% of the genome were used to evaluate the performance of the tools (**Table 9**). Performance of all tools could be increased even though TSSAR improvement was marginal. Generally, the Sp of all tools was very high (**Table 9**). The reason for that was the high number of positions where a TSS could potentially occur on the manually annotated part of the genome (e.g. 201,484 nts for *E. coli*). This leads to high numbers

of true negative annotations compared to the 181 true positive manual annotations and consequently to a high Sp. Therefore, the precision (PRC) by which a TSS was correctly annotated functioned as a metric to evaluate tool performances. The tool that performed best was TSSpredator (Sn = 0.812 and PRC = 0.498) after parameter optimization (**Table 9**). The tool with the highest PRC (0.529) was TSSAR, but its slightly higher PRC compared to TSSpredator was accompanied by a relatively low Sn (0.503). TSSer had a better Sn than TSSAR (0.757) but the lowest PRC of all tools (0.284).

Table 9. Comparison of default and optimized parameters on tool performances for detection of 181 manual annotated TSSs in *E. coli*.

Tool	Parameter	TP	FP	Sn	Sp	PRC
TSSpredator	Default	119	221	0.657	0.9989	0.350
	Optimization	147	148	**0.812**	0.9993	**0.498**
	Thomason et al., 2015	145	500	0.801	0.9975	0.225
TSSAR	Default	90	82	0.497	0.9996	0.523
	Optimization	91	81	0.503	0.9996	0.529
TSSer	Default	132	459	0.729	0.9977	0.223
	Optimization	137	346	0.757	0.9983	0.284
		Optimized tools combined				
TSSs found by all three tools		74	25	0.409	0.9999	**0.747**
TSSs found by at least one tool		162	437	**0.895**	0.9978	0.270

FP, false positive; **PRC**, precision; **Sn**, sensitivity; **Sp**, specificity; **TP**, true positive; **TSS**, transcription start site

In a next step, results of all tools were intersected leading to the highest PRC (0.747) if a TSSs was annotated by all three tools. The highest Sn (0.895) was achieved if a TSSs was considered by at least one tool, however this was accompanied by reduced PRC (0.270). The comparison showed that the optimized TSSpredator had the best performance and that the PRC could be increased by inclusion of TSSAR and TSSer results (**Table 9**). Using the optimized TSSpredator 6,241 TSSs were annotated in *E. coli* of which 3,227 were predicted with a PRC of 0.498 and for 3,015 TSSs the PRC could be increased to 0.747 by intersection with the TSSAR and TSSer results.

In order to integrate the results into published *E. coli* TSS annotations a comparison to the original study (Thomason *et al.*, 2015) and the Cappable-seq (Ettwiller *et al.*, 2016) data was performed (**Figure 11**). Of the 6,241 TSSs predicted, 5,593 (89.6%) were also

found in the original study (**Figure 11**). This was expected since the same data set was used. However, PRC of annotations was lower in the original study (**Table 9**), resulting in 14,868 predicted TSSs (**Figure 11**). This would imply three and a half times more TSSs than genes encoded on the *E. coli* genome. Given that the Sn of the presented approach and the original study was similar but the number of falsely annotated TSSs was 3.38 times higher in the latter one (**Table 9**) the presented approach performs better. For *E. coli,* another high throughput data set using Cappable-seq (Ettwiller *et al.*, 2016) for TSS annotation was available and a comparison to the presented workflow revealed that 3,526 TSSs (56.5%) were assigned by both approaches (**Figure 11**). This is currently the highest fraction of TSSs predicted by two independent high throughput techniques. However, a drawback of the Cappable-seq analysis is that in total 16,359 TSSs were predicted (**Figure 11**) which indicates again a high number of false positive annotations. In summary the presented workflow performs well because it balances the necessary Sn with a suitable Sp of annotations.

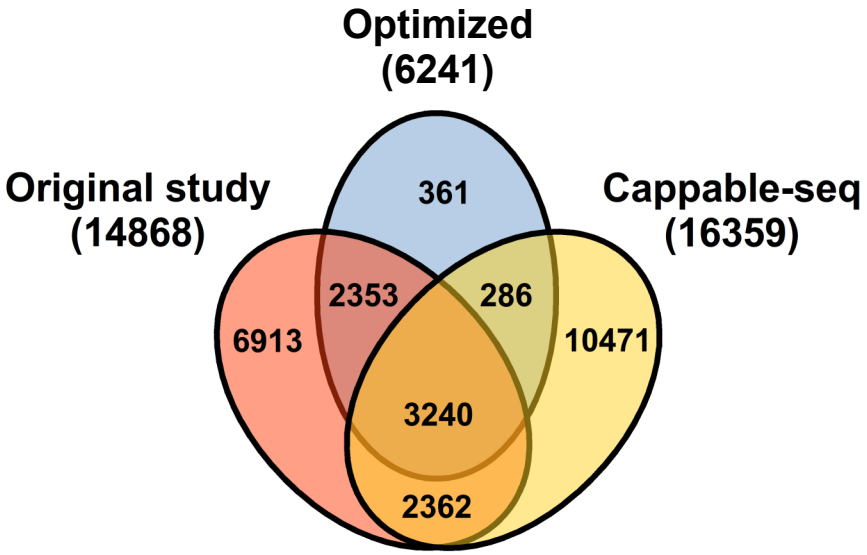


Figure 11. Venn Diagram of *E. coli* TSSs annotated by different approaches. In total 6,241 TSSs were annotated with optimized TSSpredator settings in combination with TSSAR and TSSer. Of these 5,593 (89.6%) were identified by the original dRNA-Seq analysis and 3,526 (56.5%) by an alternative Cappable-seq approach.

5.6.2 TSS annotation in *Chlamydia*

For the data sets of *C*. *abortus*, *C*. *pneumoniae*, *C*. *psittaci* and *W*. *chondrophila* the same workflow for optimization of TSSpredator, TSSAR and TSSer was applied. Similar to the analysis in *E*. *coli*, parameter optimization increased the performance of all tools (**Table 10**). The highest Sn (up to the detection of all manual annotated TSS in *C*. *pneumoniae*) was always achieved if a TSS was detected by at least one tool, however this was also always accompanied by considerably reduced PRC. In contrast, the best PRC (up to 0.833) was achieved if a TSS was detected by all three tools, but this was accompanied by a reduced Sn.

The main goal of this work was to achieve a suitable tradeoff between Sp and Sn and, as in *E*. *coli*, the tool that performed best in these terms was TSSpredator after optimization (**Table 10**). Therefore, TSSpredator results were used and in total 1,127 TSSs were annotated in *C*. *psittaci*, 888 in *C*. *abortus* and 547 in *C*. *pneumoniae*. Moreover, as in *E*. *coli*, intersecting TSSpredator with TSSAR and TSSer results lead to TSSs calls with the highest PRC (**Table 10**). In total 349 TSSs in *C*. *psittaci*, 277 in *C*. *abortus* and 79 in *C*. *pneumoniae* were annotated by such intersection leading to high PRC of 0.760, 0.737 and 0.750, respectively. The lower number of TSS detected in *C*. *pneumoniae* was due to the considerably lower sequencing depth (**Supplementary Table 1**). For *W*. *chondrophila* the highest PRC (0.875) was achieved by optimized TSSpredator alone resulting in 348 TSSs (**Table 10**). Because of the relatively low number of TSSs detected by TSSpredator alone, TSSs found by at least two tools and a PRC of 0.462 were included resulting in 999 annotated TSSs. This example emphasizes the usage of several tools for robust TSS prediction, because performance of individual tools may vary among data sets.

Table 10. Performance comparison of default and optimized settings and combination of tools.

	Tool	Parameter	TP	FP	Sn	Sp	PRC
***C. psittaci* (40)**	TSSpredator	Default	9	6	0.225	0.99988	0.600
		Optimization	35	52	**0.875**	**0.99896**	**0.402**
	TSSAR	Default	23	147	0.575	0.99832	0.135
		Optimization	24	122	0.600	0.99826	0.164
	TSSer	Default	23	84	0.575	0.99706	0.177
		Optimization	24	87	0.600	0.99756	0.216
			Optimized tools combined				
	TSSs found by all three tools		19	6	0.475	0.99988	**0.760**
	TSSs found by at least one tool		39	181	**0.975**	0.99638	0.177
***C. abortus* (45)**	**Tool**	**Parameter**	**TP**	**FP**	**Sn**	**Sp**	**PRC**
	TSSpredator	Default	4	2	0.085	0.99996	0.667
		Optimization	20	20	**0.444**	**0.99960**	**0.500**
	TSSAR	Default	25	137	0.556	0.99844	0.154
		Optimization	25	121	0.556	0.99880	0.171
	TSSer	Default	16	78	0.356	0.99726	0.170
		Optimization	16	60	0.356	0.99758	0.211
			Optimized tools combined				
	TSSs found by all three tools		14	5	0.311	0,31111	**0.737**
	TSSs found by at least one tool		44	151	**0.978**	0,99698	0.226
***C. pneumoniae* (20)**	**Tool**	**Parameter**	**TP**	**FP**	**Sn**	**Sp**	**PRC**
	TSSpredator	Default	7	9	0.350	0.99994	0.438
		Optimization	16	26	**0.800**	**0.99981**	**0.381**
	TSSAR	Default	8	13	0.400	0.99876	0.381
		Optimization	10	14	0.500	0.99884	0.417
	TSSer	Default	15	175	0.600	0.99991	0.080
		Optimization	15	161	0.750	0.99990	0.085
			Optimized tools combined				
	TSSs found by all three tools		3	1	0.150	0.99999	**0.750**
	TSSs found by at least one tool		20	179	**1.000**	0.99871	0.101
***W. chondrophila* (68)**	**Tool**	**Parameter**	**TP**	**FP**	**Sn**	**Sp**	**PRC**
	TSSpredator	Default	3	2	0.044	0.99996	0.600
		Optimization	7	1	**0.103**	**0.99998**	**0.875**
	TSSAR	Default	13	42	0.191	0.99868	0.236
		Optimization	15	38	0.221	0.99874	0.283
	TSSer	Default	27	66	0.397	0.99916	0.290
		Optimization	27	63	0.397	0.99924	0.300
			Optimized tools combined				
	TSSs found by all three tools		5	1	0.074	0.99998	**0.833**
	TSSs found by at least one tool		31	73	**0.456**	0.99854	0.298

Number next to the species in brackets are the manual annotated TSSs used for optimization. For *C. pneumoniae* only 20 TSSs were manually annotated because of the lower sequencing depth.

FP, false positive; **PRC**, precision; **Sn**, sensitivity; **Sp**, specificity; **TP**, true positive; **TSS**, transcription start site

5.6.3 Relative position of TSSs and 5'UTR length

The TSSs were classified according to their relative positions to genes in five categories (Bischler *et al.*, 2015) i.e. primary (p), secondary (s), internal (i), antisense (a), and orphan (o) (**Figure 12**). Thereby pTSS and sTSS correspond to annotated genes, whereas aTSSs and iTSSs might relate to small non-coding and antisense RNAs, respectively. These are frequently found in bacteria and have been described for *C. trachomatis* (Albrecht *et al.*, 2010) and *C. pneumoniae* (Albrecht *et al.*, 2011).

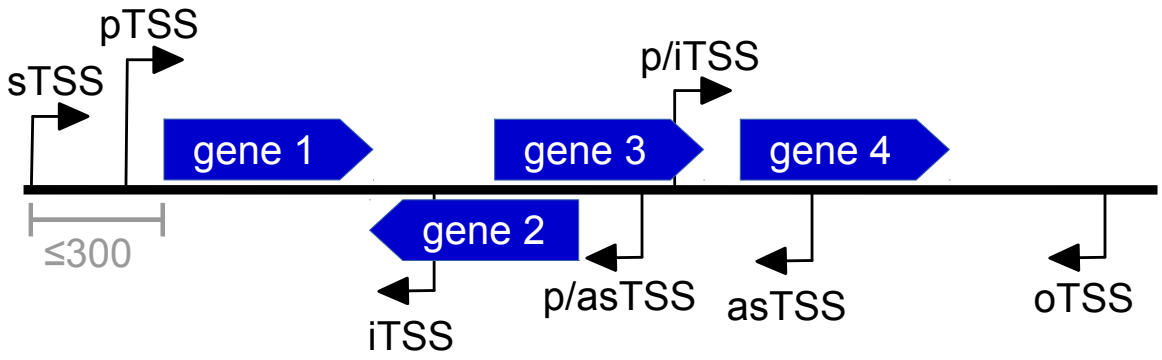


Figure 12. TSSpredator transcription start site classification. The location relative to annotated genes is depicted for the five different TSS classes i.e., primary (p), secondary (s), internal (i), antisense (a), and orphan (o). The figure was adapted from (Bischler *et al.*, 2015).

A comparison of TSSs positions (**Figure 13**) revealed that fractions of pTSSs (~26%), sTSSs (~7%), asTSSs (~16%) and oTSSs (~4%) are similar in all species. However, iTSSs differed between *E. coli* (**Figure 13A**) and *Chlamydia* (**Figure 13B-E**) given that in *E. coli* iTSSs accounted for 30.8% and in *Chlamydia* for 48.5% to 56.8% (q <0.01). One possible reason for the higher number of iTSS in *Chlamydia* might be internally encoded sRNAs, which were described previously in *C. pneumoniae* (Albrecht *et al.*, 2011). Another possibility might be the annotation of chlamydial genomes, which is mainly based on automatic pipelines using sequence comparisons.

At least some iTSSs near the predicted start codon might originate from transcripts with alternative shorter ORFs. Therefore, alternative shorter ORFs consistent with the TSS were searched up to 60 nts downstream of the translational start site. This revealed that 111 (2.3%) genes in *Chlamydia* have to be re-annotated.

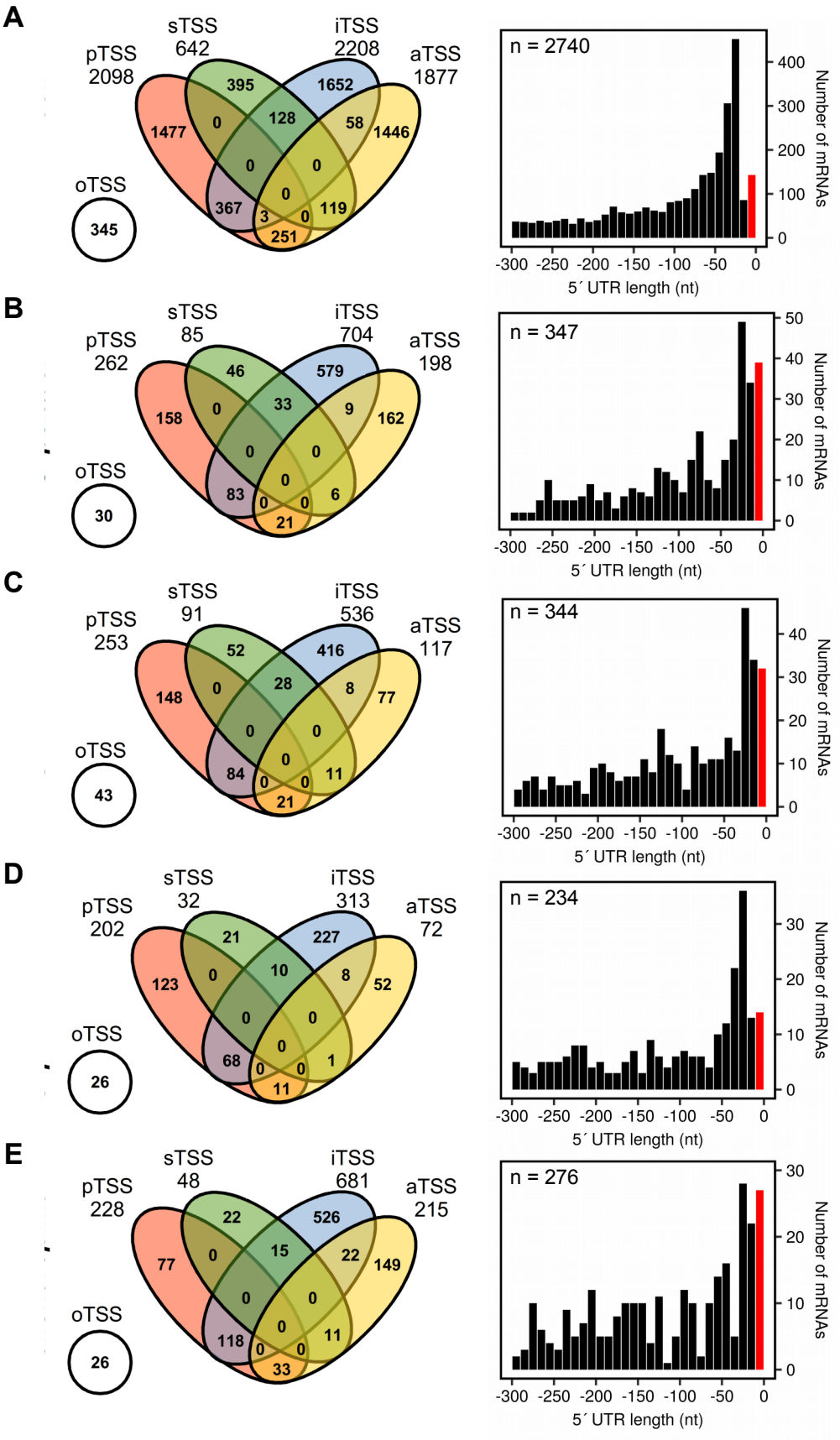

A
pTSS 2098
sTSS 642
iTSS 2208
aTSS 1877
oTSS 345
1477
395
1652
1446
128
58
367
119
251
n = 2740
Number of mRNAs
5´ UTR length (nt)
B
pTSS 262
sTSS 85
iTSS 704
aTSS 198
oTSS 30
158
46
579
162
33
9
83
6
21
n = 347
Number of mRNAs
5´ UTR length (nt)
C
pTSS 253
sTSS 91
iTSS 536
aTSS 117
oTSS 43
148
52
416
77
28
8
84
11
21
n = 344
Number of mRNAs
5´ UTR length (nt)
D
pTSS 202
sTSS 32
iTSS 313
aTSS 72
oTSS 26
123
21
227
52
10
8
68
1
11
n = 234
Number of mRNAs
5´ UTR length (nt)
E
pTSS 228
sTSS 48
iTSS 681
aTSS 215
oTSS 26
77
22
526
149
15
22
118
11
33
n = 276
Number of mRNAs
5´ UTR length (nt)

Figure 13. Position of TSSs relative to annotated genes and 5'UTR length. Categorization of TSSs according to the annotation scheme in **Figure 12** for *E. coli* (A), *C. psittaci* (B), *C. abortus* (C), *C. pneumoniae* (D) and *W. chondrophila* (E) predicted TSSs (left panel). On the right panel distribution and frequency of the 5′UTR lengths are shown. Leaderless mRNAs (5'UTR <10nt) are highlighted in red and the total number of pTSSs and sTSSs is shown in the top left of each bar graph.

A very important region of a transcript is the 5’UTR to which complementary asRNAs or sRNA can bind and influence mRNA stability and translation (reviewed in Oliva, Sahr and Buchrieser, 2015). Based on TSS annotations, the length of 5’UTRs from primary and secondary transcripts were assigned (**Figure 13**, right panel). Generally, the mean 5’UTR length was similar for all five bacteria (89 – 111 nts) and most ORFs have 5’UTRs ranging from 10 to 50 nts in length (**Figure 13**, right panel). However, also longer 5’UTRs up to 300 nts were found in all investigated bacteria. Long 5’UTRs might belong to transcripts that are more tightly regulated by sRNAs and long UTRs can present more potential complementary binding sites. Beside this, also leaderless transcripts with 5’UTRs ≤ 10 nts were found (**Figure 13**, right panel). These results are in agreement with previous studies which found leaderless mRNAs in various bacteria (Sharma *et al.*, 2010; de Groot *et al.*, 2014; Kopf *et al.*, 2014; Ettwiller *et al.*, 2016) including *C. pneumoniae* (Albrecht *et al.*, 2011) indicating that these are common.

5.6.4 Conservation of chlamydial TSSs and regulatory regions

Chlamydial ORFs are highly conserved and this motivated the analysis of TSS conservation among *Chlamydia* for comparison. In order to analyze the conservation of TSSs, homologous genes were searched for pTSSs as well as sTSSs and relative positions were compared. The close relatives *C. psittaci* and *C. abortus* shared the highest number of conserved TSSs and of the 599 TSSs found for homologous genes 230 (38.4%) had the same relative position (±3 nts). As *C. pneumoniae* was included in the comparison 72 homologous genes were found for which pTSSs or sTSSs were annotated in all three

species. However, only 14 TSSs (19.4%) with the same relative position in all three species were found of which one is a conserved TSS of *ompA*. After inclusion of *W. chondrophila* only for seven homologous genes a pTSS or sTSS was found in all four chlamydial species. For these seven genes no TSS had the same relative position (±3 nts) in all *Chlamydia*. In conclusion, the comparison showed a high level of diversity in the relative positions of TSSs and consequently 5'UTR lengths of homologous genes.

After evaluating the conservation of TSS positions corresponding ORF, 5'UTR and promoter sequences were compared (**Figure 14**). For this, the sequences were extracted based on TSSs found in homologous chlamydial genes. In total, 76 genes in *C. psittaci*, *C. abortus* and *C. pneumoniae* were found for which also a TSS in all three organisms was annotated. *W. chondrophila* had to be excluded from the comparison because of the 558 conserved ORFs only for 14 on average a corresponding TSS was found in *C. psittaci*, *C. abortus* and *C. pneumoniae*. The comparison showed that ORF sequences of *C. psittaci* and *C. abortus* are highly conserved (mean = 95.0%) compared to *C. pneumoniae* (mean = 71.8%). In contrast, promoter sequences are less conserved (**Figure 14**) and two groups of promoter sequences were found i.e., highly conserved (~70% – 100% identity) and less conserved (~50% identity). In the 5'UTR sequences the conservation was lowest and especially variance of sequence identity was higher (**Figure 14**). This means that both, highly conserved and variable UTR sequences were found. These results are in agreement with a comparison between *E. coli* and *Klebsiella pneumoniae*, which revealed that the ORF was the most conserved element, followed by the promoter and the 5'UTR as the least conserved region (Kim *et al.*, 2012).

5.6.5 Conserved motifs in chlamydial promoter regions

Knowing the exact TSS positions enabled the identification of consensus sequences in chlamydial promoters. Thus far, these have been only experimentally defined by large-scale sequence comparisons for *C. pneumoniae* (Albrecht *et al.*, 2011). In order to analyze the promoters, sequences -1 to -40 nts upstream of all 3,561 chlamydial TSSs

were extracted and searched for conserved motifs. The analysis revealed 435 promoters (12.2%) with a highly conserved σ^{66} motif (p <0.001) at positions -35 to -7 and a spacer ranging from N_{15} to N_{20} (**Figure 15A**). The number of promoters harboring a σ^{66} motif could be further increased to 1,590 (44.7%) by applying a less stringent cutoff (p <0.01). Consequently, the resulting motif was slightly less conserved, but still closely resembled the *E. coli* σ^{70} motif (**Figure 15A**).

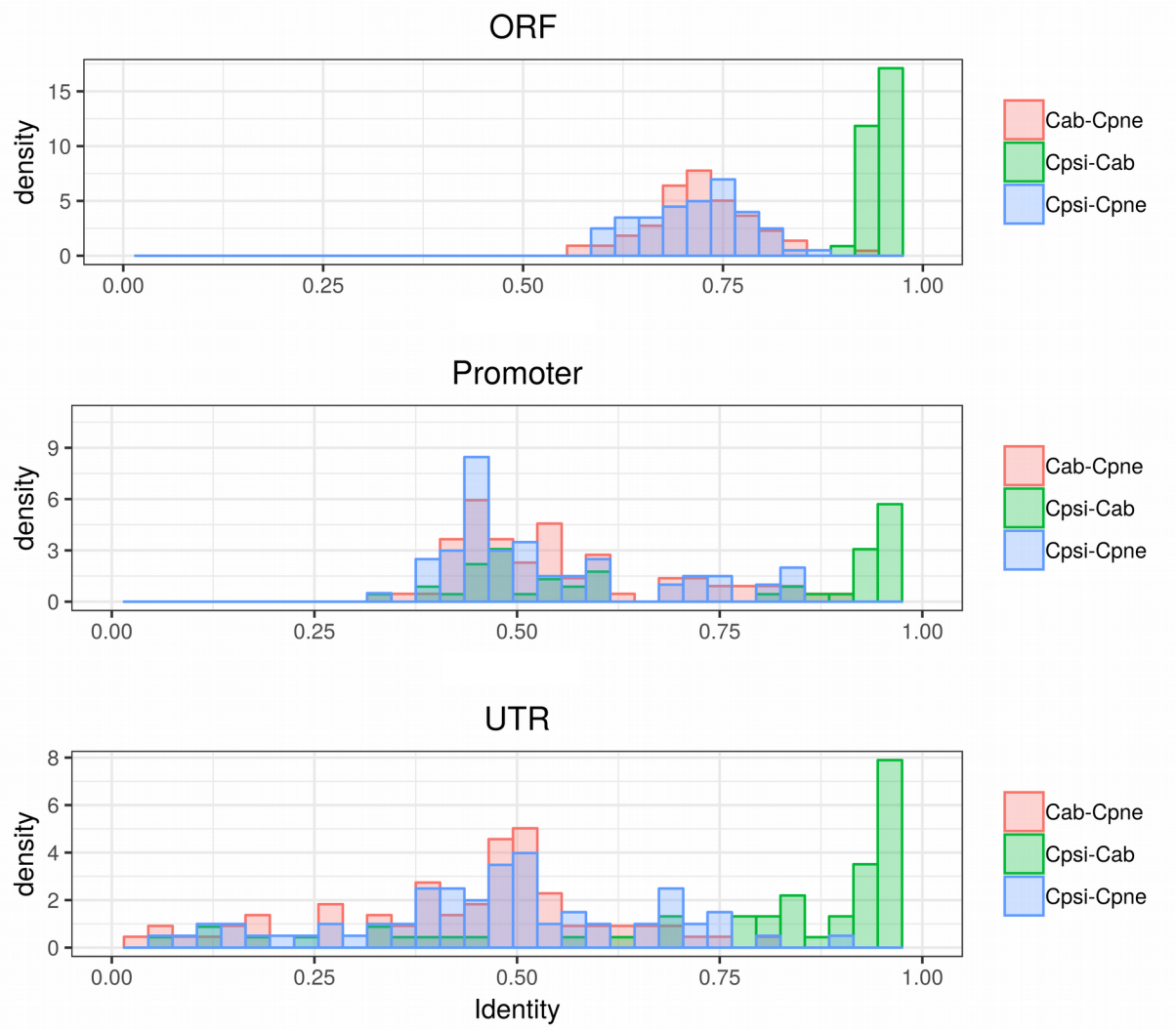


Figure 14. ORF, promoter and 5'UTR conservation of homologous genes. Density histograms showing sequence identity frequencies of 76 homologous ORFs and the corresponding 5'UTR as well as promoter sequences. Cab, *C. abortus*; Cpne, *C. pneumoniae*; Cpsi, *C. psittaci*; ORF, open reading frame; UTR, untranslated region.

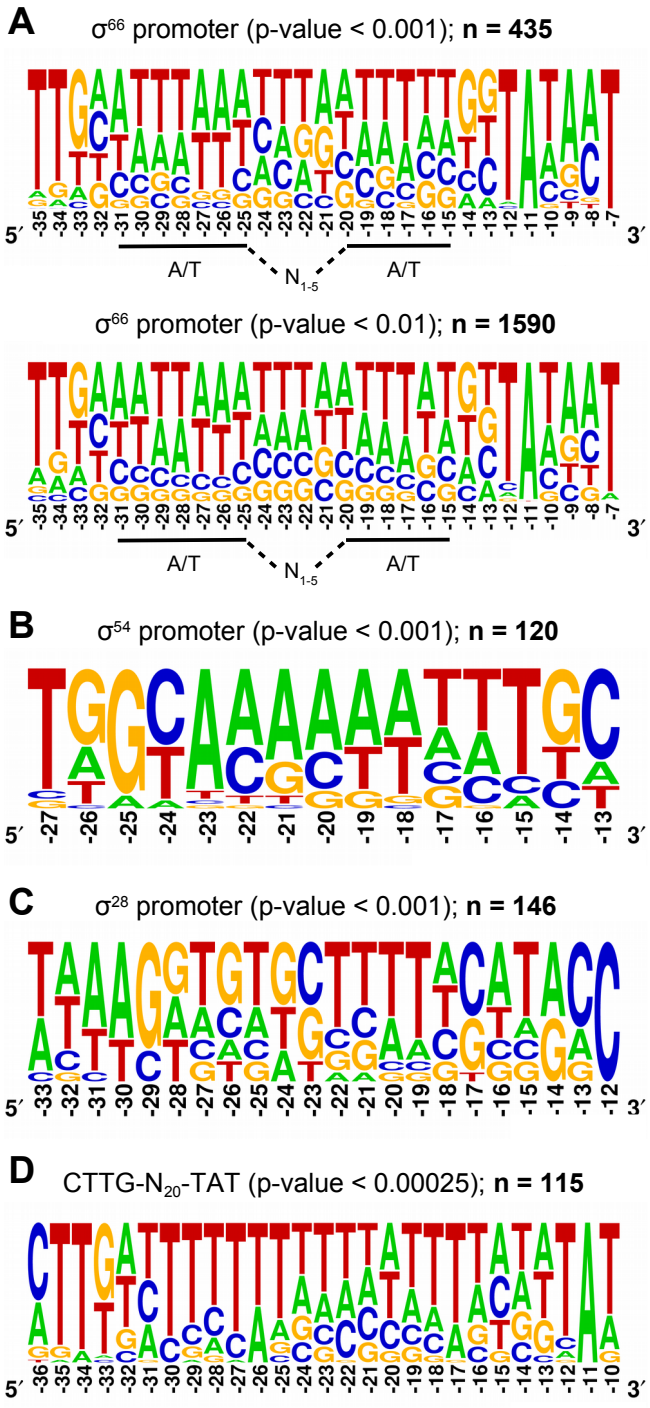


Figure 15. Motif detection in chlamydial promoters. Motif search in chlamydial promoters revealed prominent σ66 (**A**), σ54 (**B**), σ28 (**C**) and the CTTG-N_{21}-TAT (**D**) motif. Sequence motifs were generated using WebLogo (Crooks *et al.*, 2004) and x-axes represent the relative positions to the TSS.

As described elsewhere (Albrecht *et al.*, 2011), the chlamydial σ66 -35 box motif TTK is shorter than the *E. coli* σ70 consensus sequence TTGACA (**Figure 15A**). However, the -10 box resembles the *E. coli* TATAAT, but with the consensus TANNNT. In addition, two A/T rich stretches between the -10 and the -35 box were found located at positions -31 to -25 and -20 to -15 relative to the TSS. The fraction of promoters containing the σ66 motif was larger (q <0.001) than in randomly generated sequences (same nucleotide composition and length) and in addition the motif was enriched in promoters of pTSSs (q <0.001). This indicates an involvement of the σ66 motif in the expression of protein coding genes rather than regulatory RNAs.

A second known chlamydial sigma factor is σ54. In general σ54 has the most conserved consensus sequence among bacteria (Barrios, Valderrama and Morett, 1999), however in *Chlamydia* its binding motif still needs to be elucidated. Therefore, the σ54

motif of *E. coli* was searched in the chlamydial promoters, yielding 120 potential σ^{54} binding sites (p <0.001) and the consensus TDGYAM-N_5-HHBH (**Figure 15B**). Compared to the consensus sequence of other bacteria (TGGCAC-N_5-TTGC) the chlamydial -24 box appears to be more conserved than the -12 box closer to the TSS (**Figure 15B**). The number of promoters harboring the σ^{54} is smaller than the σ^{66} indicating a high specificity of the transcription factor. This is in agreement with *E. coli* for which also only ~250 σ^{54} binding sites have been estimated (Bonocora *et al.*, 2015). The third sigma factor found in *Chlamydia* is σ^{28}, which was shown to be expressed at the late stage of infection (Fahr *et al.*, 1995). Searching for the *E. coli* σ^{28} motif yielded 146 predictions (p <0.00025) in all chlamydial promoters (**Figure 15C**). This indicates a high specificity of the transcription factor (similar to σ^{54}) for which in *E. coli* also merely ~100 σ^{28} binding sites have been estimated (Yu, Kibler and Tan, 2006).

In addition to the σ^{66}, a second prominent motif in *C. pneumoniae* has been identified that derived from an alignment of 24 promoters, of which 10 belong to the Pmp family (Albrecht *et al.*, 2011). These promoters share the motif CTTG at the -35 region and TAT at the -10 box with a 21 nt T-rich spacer in between. In the chlamydial promoters 115 sequences containing the motif were found (**Figure 15D**). The fraction of promoters containing the motif is larger (q <0.001) than in randomly generated sequences with the same nucleotide composition and in addition, a higher fraction (69) was found in promoters corresponding to pTSSs (q <0.001). Moreover, another 14 Pmp genes harboring the motif were found and also 22 sequences in *W. chondrophila*. This indicates a high conservation of this novel motif since *Chlamydia* and *Waddlia* diverged about one billion years ago (de Barsy and Greub, 2013). In summary 1,971 (55.3%) promoters with conserved σ or the CTTG-N_{21}-TAT motifs were identified.

In addition to the guided search, the chlamydial promoters have been applied for *de novo* (Bailey *et al.*, 2015) motif discovery. This revealed four significant motifs within 186 promoters (**Figure 16**). Thereby, the two most abundant motifs are HDMHRAAAAAGMYKH in 87 promoters at positions -29 to -15 and the shorter KSSGATCT

at position -8 to -1 in 63 promoters (**Figure 16**). Moreover, two long motifs with 29 nts and the consensus YWYTTRTTAKMMWKTAWAAGRAGTTATYG and CCSTGWARGAGAAGWRSAARCAGGARWCG at positions -34 to -6 were found in 19 and 17 promoters, respectively (**Figure 16**).

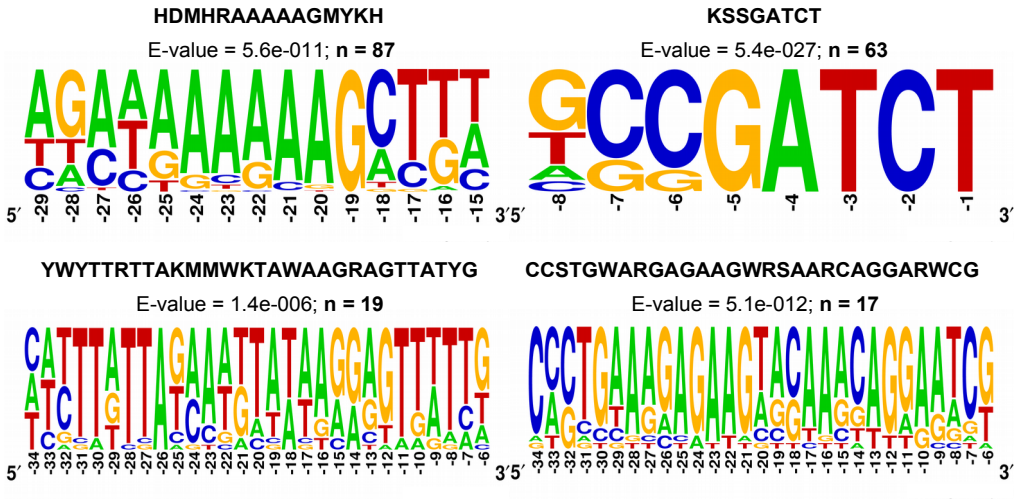


Figure 16. *De novo* motif discovery in chlamydial promoters. Chlamydial promoter sequences (positions -40 to -1) were searched for motifs using MEME yielding four enriched motifs with more than 15 occurrences. Sequence motifs were generated using WebLogo (Crooks *et al.*, 2004) and x-axes represent the relative positions to the transcription start site.

5.7 Expression of homologous genes in *C. psittaci*, *C. abortus* and *W. chondrophila*

It has been indicated that differential expression of transcripts and consequently protein levels, rather than sequence dissimilarities alone, might contribute to the differing host specificity and tissue tropism of *Chlamydia* (Braukmann *et al.*, 2012). Therefore, the very first inter-species comparison of gene expression in *Chlamydia* was performed and for that

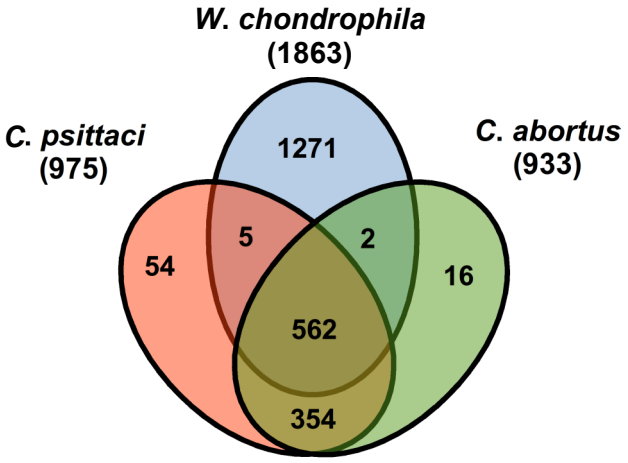


Figure 17. Comparison of homologous and species-specific protein coding genes. Venn diagram showing the number of protein-coding genes that are unique or shared among *C. psittaci*, *C. abortus* and *W. chondrophila*.

purpose homologous genes had to be identified. Altogether 562 homologs present in all three pathogens were found (**Figure 17**). These genes participate in essential cellular processes like respiration, cell division, DNA, fatty acid, protein and RNA metabolism. Potential VFs are mainly involved in cell wall synthesis, membrane transport and formation of chlamydial OMC (**Supplementary Table 4**). *C. psittaci* and *C. abortus* share the vast majority i.e., 916 CDSs. Of these 354 are not present in *W. chondrophila* (**Figure 17**) and these genes mainly encode Pmps, Incs and T3SEs (**Supplementary Table 4**), all of which are classes of proteins involved in virulence (Rockey, Lenart and Stephens, 2000). Beside these homologs, there are only few species-specific genes in *C. psittaci* and *C. abortus*, i.e., fifty-four and sixteen, respectively (**Figure 17**).

The inter-species comparison of gene expression revealed that among the 916 genes shared among *C. psittaci* and *C. abortus* (**Figure 17**), the 354 shared not with *W. chondrophila* correlate more in their expression between the close relatives (Pearson = 0.67, p-value <0.001) than the 562 core genes (Pearson = 0.49, p-value <0.001). This implies more different gene expression of core genes in *C. psittaci* and *C. abortus* that are mainly involved in essential cellular processes (**Supplementary Table 4**). Interestingly, the core gene expression was more similar (Pearson = 0.75, p-value <0.001) among *C. abortus* and *W. chondrophila,* which might be due similar metabolic requirements. Despite the differences, a core transcriptome in EBs and RBs was found that consisted of similarly expressed genes in all three pathogens. The number of these core transcriptome genes were 209 (37%) and 327 (58%), respectively. Differential expression of homologous genes, especially VFs, might be of great relevance in the *Chlamydia* disease outcome (Braukmann *et al.*, 2012). Therefore, the expression of 30 known VFs and immunogenic genes shared between *C. psittaci* and *C. abortus* is summarized in **Figure 18**. To confirm the RNA-Seq results, RT-qPCR was used to validate the expression of *cap1*, *copB_1*, *incA*, *ompA* and *sinC* (**Figure 18A**). Similar to the intra-species comparison, a high correlation between RT-qPCR and RNA-Seq results (0.98) was found in the inter-species comparison.

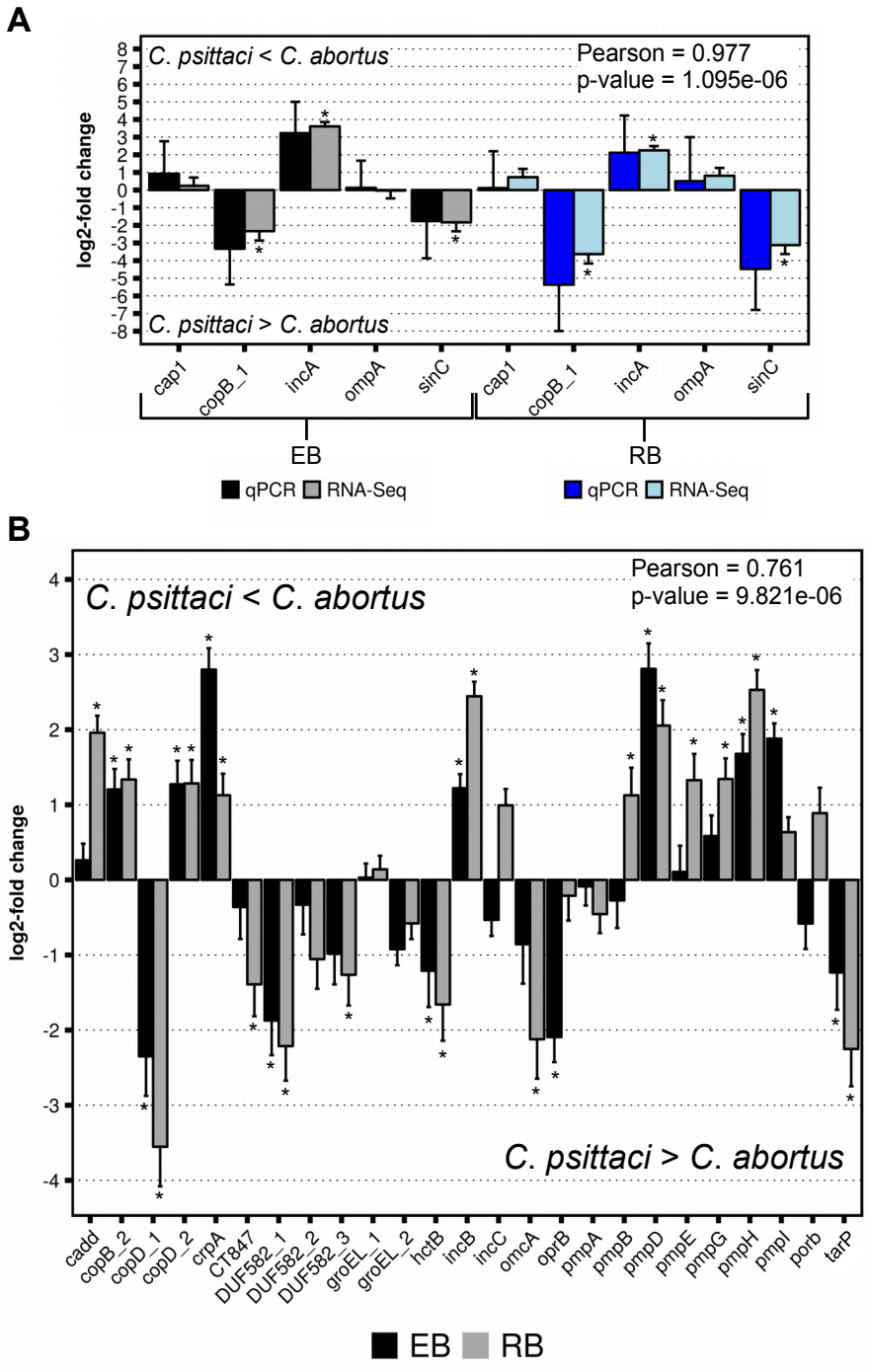
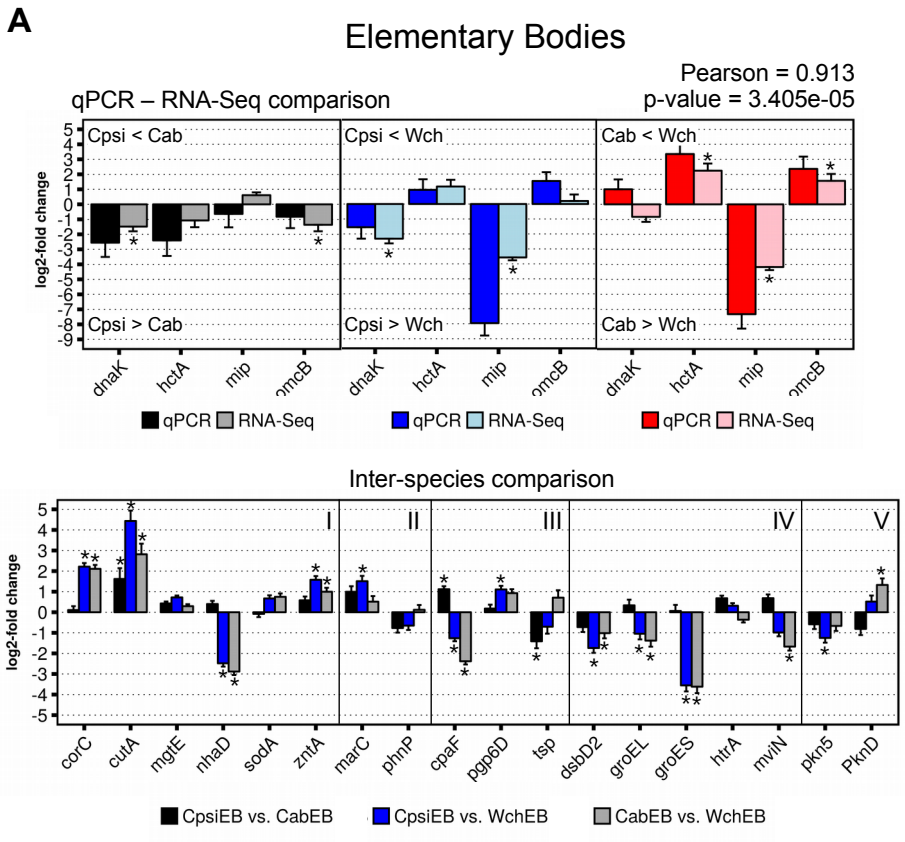


Figure 18. Expression of homologous virulence factors in *C. psittaci* and *C. abortus*. The log2-fold changes of expression from 30 virulence and immunogenic factors present in *C. psittaci* and *C. abortus* but absent in *W. chondrophila* are shown. (A) Confirmation of the RNA-Seq results with RT-qPCR. (B) Differential expression of virulence factors correlates in EBs and RBs of *C. psittaci* and *C. abortus*. Genes for which no gene name was found are specified with the *C. trachomatis* NCBI locus tag. Significance level is depicted by asterisk (*) and represents BH adjusted p-value <0.01 and an absolute log2-fold change >1.0. Standard error of log2-fold changes are indicated.


A
Elementary Bodies
Pearson = 0.913
p-value = 3.405e-05
qPCR – RNA-Seq comparison
Cpsi < Cab
Cpsi < Wch
Cab < Wch
Cpsi > Cab
Cpsi > Wch
Cab > Wch
log2-fold change
dnaK
hctA
mip
omcB
qPCR
RNA-Seq
Inter-species comparison
I
II
III
IV
V
corC
cutA
mgtE
nhaD
sodA
zntA
marC
phnP
cpaF
pgp6D
tsp
dsbD2
groEL
groES
htrA
mviN
pkn5
PknD
CpsiEB vs. CabEB
CpsiEB vs. WchEB
CabEB vs. WchEB

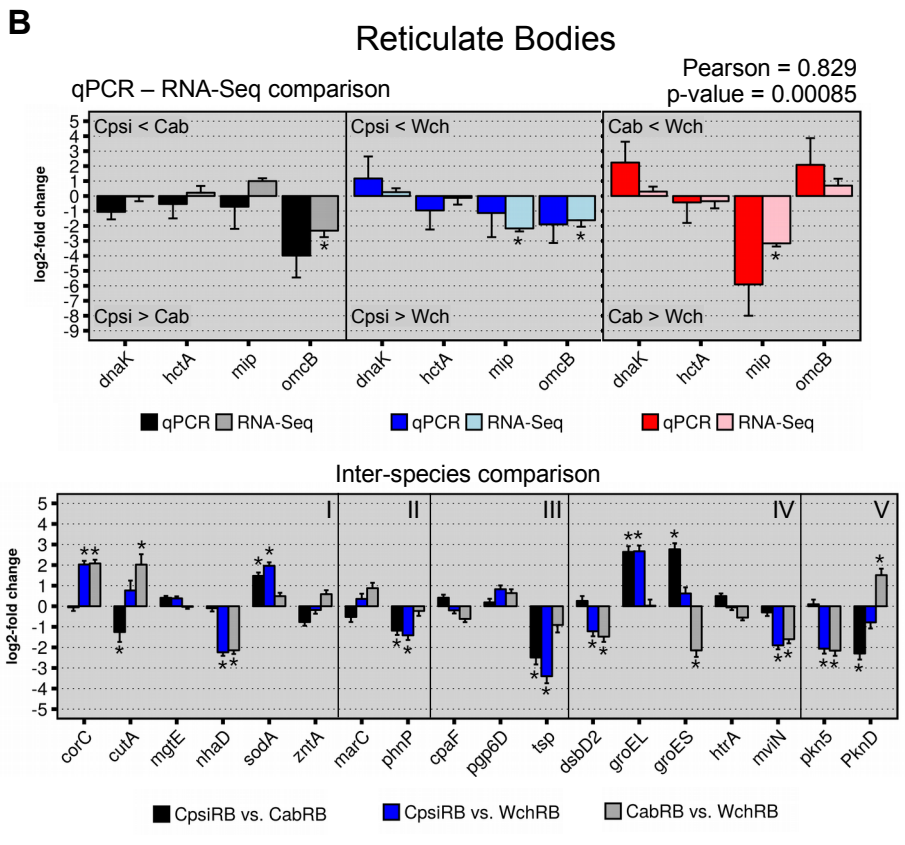

B
Reticulate Bodies
Pearson = 0.829
p-value = 0.00085
qPCR – RNA-Seq comparison
Cpsi < Cab
Cpsi < Wch
Cab < Wch
Cpsi > Cab
Cpsi > Wch
Cab > Wch
log2-fold change
dnaK
hctA
mip
omcB
qPCR
RNA-Seq
Inter-species comparison
I
II
III
IV
V
corC
cutA
mgtE
nhaD
sodA
zntA
marC
phnP
cpaF
pgp6D
tsp
dsbD2
groEL
groES
htrA
mviN
pkn5
PknD
CpsiRB vs. CabRB
CpsiRB vs. WchRB
CabRB vs. WchRB

Figure 19. Expression of homologous virulence factors in *C. psittaci*, *C. abortus* and *W. chondrophila*. The log2-fold changes of expression from 22 virulence and immunogenic factors present in *C. psittaci*, C. *abortus* and *W. chondrophila* are shown. (A) Confirmation of the RNA-Seq results in EBs with RT-qPCR (upper panel) and inter-species comparison (lower panel). Genes are sorted according to their functions: (I) stress response, (II) drug resistance, (III) manipulation of host cell immune response, (IV) diverse functions and (V) effectors secreted by the type III secretion system. (B) Confirmation of the RNA-Seq results with RT-qPCR in RBs (upper panel) and inter-species comparison (lower panel). Significance level is depicted by asterisk (*) and represents Benjamini-Hochberg adjusted p-value <0.01 and an absolute log2-fold change >1.0. Standard error of log2-fold changes are indicated.

Interestingly, most of the VFs are differentially expressed in *C. psittaci* and *C. abortus* (**Figure 18A**) and the log2-fold changes between EBs and RBs correlate (0.76). An example for a VF higher expressed in *C. psittaci* EBs and RBs is *sinC*. The protein is conserved in *C. psittaci*, *C. abortus*, *C. caviae* and *C. felis*, but not in the major human pathogens *C. trachomatis* and *C. pneumoniae*. SINC is secreted and targets the inner nuclear membrane of the host and also of uninfected neighbors (Mojica *et al.*, 2015). Other VFs more abundant in *C. psittaci* were *hctB* and *copD_1*, whereas *pmpD*, *pmpH*, *incA* and *incB* were higher expressed in *C. abortus* (**Figure 18B**). In total, 562 homologous genes (**Figure 17**) present in all three pathogens were identified, which mainly participate in essential cellular processes (**Supplementary Table 4**). However, among these core genes are 22 VFs and immunogenic factors. Four of them (*dnaK*, *hctA*, *mip* and *omcB*) were used for confirmation of the RNA-Seq results with RT-qPCR in inter-species comparisons (**Figure 19**, upper panels).

Again, the Pearson correlation coefficient of 0.913 in EBs and 0.829 in RBs demonstrated a high degree of consensus between the RT-qPCR and RNA-Seq results. Interestingly, as in the comparison of *C. psittaci* and *C. abortus* most of the homologous VFs are differentially (e.g. *hctA* and *omcB*) and only few similarly expressed (e.g. *hrtA*) in all species (**Figure 19**). In general, the expression of these 22 VFs is more diverse in EBs and RBs than the 30 VFs exclusively shared among *C. psittaci* and *C. abortus* (**Figure 19B**).

6 Discussion

6.1 Purification of *Chlamydia* using iodixanol preserves integrity of EBs and RBs

The present work focused on analyzing the transcriptomic differences of EBs and RBs. Consequently, it was an important preparatory work to purify these developmental forms from infected human cells. Available protocols for EB and RB purification are based on density gradient centrifugation to separate the dense EBs (1.21 g cm^{-3}) and less dense RBs (1.14 – 1.18 g cm^{-3}). However, the existing protocols seemed unsuitable for RNA-Seq experiments because the usage of non-inert sucrose for gradient preparation (Tamura and Higashi, 1963; Bose and Paul, 1982; Albrecht et al., 2011), or they required long implementation time due to isopycnic centrifugation or consecutive ultracentrifugation steps (Tamura, Matsumoto and Higashi, 1967; Friis, 1972; Howard, Orenstein and King, 1974; Caldwell, Kromhout and Schachter, 1981; Mukhopadhyay *et al.*, 2004; Wehrl *et al.*, 2004; Marques *et al.*, 2010). A promising medium for gradient preparation was iodixanol, a non-toxic, inert, non-ionic and isoosmotic medium that has been widely applied in organelle (Graham, Ford and Rickwood, 1994), virus (Lindenbach *et al.*, 2005) or protein purification (Yee *et al.*, 2008). The established protocol has the following advantages: (I) avoiding high MOIs to mimic more natural infection conditions, (II) feasibility of the procedure at low temperatures (4°C), (III) fast mechanical disruption of host cells, (IV) considerably faster execution time (4 h) and (VI) suitability to purify RBs and EBs from different chlamydial species. To evaluate the protocol performance, TEM imaging, RNA integrity analyses and reinfection assays were performed. After purification, *Chlamydia* were enriched and RB fractions were virtually free of EBs and *vice versa* (**Figure 3**). Some host derived impurities were still present but these are almost inevitable due to similar densities of mitochondria (1.13 g cm^{-3} – 1.16 g cm^{-3}) as well as smooth and rough endoplasmatic reticulum fragments (1.15 g cm^{-3} – 1.19 g cm^{-3}) to chlamydial RBs and

EBs (Friis, 1972). More important was that the RNA from both, purified RB and EB fractions were highly enriched for prokaryotic rRNA thus revealing successful depletion of host transcripts (**Figure 4**). These would have made up the majority of transcripts in the libraries if they were not depleted. Moreover, reinfection assays demonstrated that EB fractions were >100 times more infectious than RB fractions (**Table 7**). Since TEM imaging could not be performed for each of the 40 replicates considered for RNA-Seq the reinfection assays in combination with the high quality RNA demonstrated the purity of isolated EB and RB fractions.

In summary, the developed protocol showed for the first time that iodixanol is an excellent gradient medium for purification of intracellular bacteria for experiments like RNA-Seq. Furthermore, the protocol is applicable to other chlamydial species right away because of similar densities of RBs and EBs within *Chlamydiaceae* (Friis, 1972; Howard, Orenstein and King, 1974). Because of the stated advantages of the developed protocol for EB and RB purification it could be published as part of the presented work in the Journal of Microbiological Methods (Beder *et al.*, 2016).

6.2 Gene expression analysis in EBs and RBs

Generally, the RNA-Seq data was of high quality (**Figure 6**) and alignment results showed only small fractions of reads not mapping to chlamydial or host genomes (**Figure 7**). Due to the fact that numbers of mRNA copies from individual genes can differ by several orders of magnitude an adequate sequencing depth is required to detect transcripts with low copy numbers. To increase the sequencing depth rRNAs were selectively removed, which resulted in ~49% of reads on average mapping to mRNAs, antisense RNAs and to intergenic regions (**Figure 8**). This means that the rRNA fraction, which usually represents 95% – 97% in prokaryotes (Rosenow, 2001) was highly reduced by the Ribo-Zero™ treatment. Consequently, high coverages were achieved (**Supplementary Table 1**) and virtually all genes detected to be expressed. In the next step, primary transcripts were enriched using TEX, an enzyme that selectively degrades

rRNAs and mRNAs harboring a 5‘P end, which designates them for decay (Celesnik, Deana and Belasco, 2007). Consequently, the fraction of reads mapping to rRNAs was significantly decreased (**Figure 8**) and primary transcripts enriched for reads towards the TSS in TEX treated samples (**Figure 2**).

Initially the incorporation of the *C*. *trachomatis* and *C*. *pneumoniae* dRNA-Seq data was planned (Albrecht *et al.*, 2010, 2011), but reanalysis with DESeq2 resulted in no significant results because of the low sequencing depth and small number of replicates. Therefore, the presented DEG analysis focused on EB and RB transcriptomes of *C*. *psittaci*, *C*. *abortus* and *W*. *chondrophila* only. Generally, EB and RB transcriptomes are very different and the number of DEGs sum up to 63%, 34% and 56% of all genes in *C*. *psittaci*, *C*. *abortus* and *W*. *chondrophila*, respectively (**Table 8**). In order to transfer biological functions to the large number of DEGs, a categorization of genes according to RAST database, by T3SE prediction and assembly of Inc, Pmp and PZ genes from the literature was performed. Differential expressed genes within the 27 categories were tested for over- and under-representation. This revealed that categories like “Amino Acids and Derivates”, “Cofactors, Vitamins, Prosthetic, Groups, Pigments” or “RNA Metabolism” are overrepresented RBs, whereas genes involved in “DNA Metabolism” are underrepresented in EBs (**Supplementary Table 2**). This is in accordance to the biological functions of the replicative RBs and infectious EBs. However, in general over- and under-representation of genes in the categories were heterogeneous among the three species indicating that different biological processes are emphasized. Interestingly, categories enriched in *W*. *chondrophila* EBs are “Carbohydrates” and “Respiration” (**Supplementary Table 2**), which indicates respiratory activity in the infectious *W*. *chondrophila* state, as previously demonstrated for another environmental *Chlamydia* (Sixt *et al.*, 2013).

The DEGs also included numerous VFs. Up-regulated VFs in EBs are the chaperones *dnaK*, *groEL* and *groES*, which are responsible for preventing misfolding and aggregation of proteins (**Supplementary Table 3**). During differentiation into the persistent forms

increased chaperone activity might be required to keep proteins within the cytoplasm in an inactive storage state for immediate activation after infection. In concordance these chaperones were also found to be late expressed genes in *C. trachomatis* and *C. pneumoniae* (Belland, Zhong, *et al.*, 2003; Mäurer *et al.*, 2007). Another class of chaperones involved in the T3S are known to be enriched in EBs (Saka *et al.*, 2011) and were also found to be transcriptional up-regulated in the infectious state. The up-regulation of these T3S chaperones in the infectious state fits well, because most proteins are preloaded in an inactive form in EBs (reviewed in Abdelrahman and Belland, 2005; Jewett *et al.*, 2010). The detailed comparison of gene expression in chlamydial EBs and RBs could be published as part of the presented work in BMC Genomics (Beder and Saluz, 2018).

Another interesting feature observed for *Chlamydia* is that at early stages of infection the EBs contain mRNAs carried over from the previous developmental cycle (Hatch, Miceli and Sublett, 1986). These carryover transcripts may not result in protein synthesis, but degrade early after infection (Hatch, Miceli and Sublett, 1986; Humphrys *et al.*, 2013). Because TEX digests mature mRNAs, down-regulated genes in EBs might to some extent represent these carryover transcripts, whereas up-regulation indicates new expression. TEX enriched transcripts might also be preloaded in EBs in order to accelerate corresponding protein synthesis. In total 161 and 104 transcripts were enriched upon TEX treatment in EBs of *C. psittaci* and *C. abortus*, respectively (**Table 8**). These can either represent very late or very early synthesized mRNAs. Interestingly, an overlap of 62 mRNAs enriched in both species was found, which is significantly more than expected by chance (p-value <2.2e-16). These 62 genes included six Incs and two T3SEs, which are promising candidates for VFs involved in infection of both species. Specific for *C. psittaci* was that three genes from the PZ were enriched by TEX of which two are not shared with *C. abortus*. The PZ is a major source of deviation from sequence conservation (Knittler *et al.*, 2014). Consequently, the PZ is assumed to harbor key genomic features for species-specific adaptation (Thomson *et al.*, 2005). Contrary, in *C. abortus* EBs *incA* was found to

be enriched upon TEX treatment and this might add to specific traits of *C. abortus*. The highest number of TEX enriched transcripts (287) was found in *W. chondrophila* EBs, but this was probably due to the twofold higher number of CDSs. However, compared to *C. psittaci* and *C. abortus* only 18 and 10 overlapping homologs were found, respectively. This shows that TEX enriches a similar set of genes in EBs of *C. psittaci* and *C. abortus*, however not in *W. chondrophila*.

6.3 TSS annotation and conservation of cis-regulatory elements

In order to reliable annotate TSS, individual tools were optimized in terms of Sn and PRC (**Table 9**, **Table 10**). Generally, the performance of all tools could be increased by optimizing the parameters and the tool with the best overall performance was TSSpredator. Moreover, PRC of annotations could be further increased by intersection with TSSAR and TSSer results. The highest Sn was always achieved if a TSS was detected by at least one tool, however this was also always accompanied by considerably reduced PRC (**Table 9**, **Table 10**). In contrast, the best PRC was achieved if a TSS was detected by all three tools, but this was accompanied by a reduced Sn. Since the main goal of this work was to achieve a suitable tradeoff between Sp and Sn, optimized TSSpredator results were used and intersected with TSSAR and TSSer annotations. In total 9,802 TSSs were annotated in *E. coli* and the *Chlamydia* with an average Sp of 0.9996 and PRC of 0.588. Because of the high Sp and PRC of TSS annotations the presented workflow performed very well which was demonstrated by the reanalysis of the *E. coli* dRNA-Seq data. The comparison to the original study (Thomason *et al.*, 2015) and Cappable-Seq technique (Ettwiller *et al.*, 2016) showed that both published approaches annotated numbers of TSSs >3.3 times excessing the number of genes (**Figure 11**). Both studies together would suggest 22,986 TSSs which would mean the presence of 5.5 times more individual transcripts than annotated genes. This indicates a high fraction of false positive annotations by both analyses and the presented workflow overcomes this limitation with its high PRC. Moreover, 3,526 TSSs (56.5%) were assigned by both, the

reanalysis of dRNA-Seq data and the Cappable-Seq approach (**Figure 11**). This is currently the highest fraction of TSSs predicted by two independent high throughput techniques.

In summary, the reanalysis of the dRNA-Seq data of *E. coli* (Thomason *et al.,* 2015) using different tools followed by validation demonstrated the suitability of the presented workflow and a TSS annotation map with unique performance was generated for the model organism. In *Chlamydia* the same workflow resulted in 1,127 annotated TSSs in *C. psittaci*, 888 in *C. abortus*, 547 in *C. pneumoniae and* 999 in *W. chondrophila*. The lower number of TSSs annotated in *C. pneumoniae* was to the lower sequencing depth (**Supplementary Table 1**). A comparison of TSSs positions (**Figure 13**) showed that fractions of pTSSs, sTSSs, asTSSs and oTSSs are similar in all species, however, iTSSs differed between *E. coli* and *Chlamydia*. One possible reason for the higher number of iTSS in *Chlamydia* might be the annotation of chlamydial genomes, which is mainly based on automatic pipelines using sequence comparisons. Therefore, at least some iTSSs near the predicted start codon might originate from transcripts with alternative shorter ORFs. Indeed, 111 genes in *Chlamydia* were found to have shorter ORFs consistent with the annotated TSSs and these genes have to be re-annotated.

An important hypothesis in the presented work was the usage of developmental stage-specific promoters in EBs and RBs, which consequently would require different utilization of TSSs. The presence of these stage-specific promoters has been demonstrated previously for the *C. trachomatis* plasmid gene pL2-02 (Ricci *et al.*, 1993; Albrecht *et al.*, 2010) and was indicated for four genes of *C. pneumoniae* (*rpsA*, CPn0365, *fabI* and CPn0408) (Albrecht *et al.*, 2011). The resulting transcripts may have varying 5'UTR lengths, which could contribute to the different expression levels in both developmental forms. Of the 1,201 chlamydial pTSSs and sTSSs only 12 were found to be different in EBs and RBs. This indicates that stage specific promoters are not a general feature responsible for differential gene expression in EBs and RBs but may exist for few selected ORFs only. Therefore, differential gene expression in EBs and RBs might be rather

controlled by other mechanisms like transcription factor accessibility of DNA or post-transcriptional regulation on the RNA levels.

Using the annotations it was now possible to determine the relative positions of each TSS and the length of corresponding 5'UTRs (**Figure 13**). The 5'UTRs regulate gene expression on both, transcriptional and translational levels by processes determining for example mRNA stability (reviewed in Van Assche *et al.*, 2015). Generally, the mean 5'UTR length was similar for all five bacteria (89 – 111 nts) and most ORFs have 5'UTRs ranging from 10 to 50 nts in length (**Figure 13**, right panel). However, also longer 5'UTRs up to 300 nts were found in all investigated bacteria. Long 5'UTRs might belong to transcripts that are more tightly regulated by sRNAs and present more potential complementary binding sites. In contrast, also leaderless transcripts with 5'UTRs ≤ 10 nts were found (**Figure 13**, right panel). These results are in agreement with previous studies which found leaderless mRNAs in various bacteria (Sharma *et al.*, 2010; de Groot *et al.*, 2014; Kopf *et al.*, 2014; Ettwiller *et al.*, 2016) including *C. pneumoniae* (Albrecht *et al.*, 2011). This indicates that the presence of leaderless transcripts is a common feature in bacteria. Interestingly, the 5'UTR length does not correlate with the expression levels of the transcripts showing that there is no simple association between these two metrics.

An important chlamydial VF with a long 5'UTR that is highly expressed is the major outer membrane protein *ompA*. OmpA constitutes more than 60% of the total outer membrane protein content in *Chlamydia* (Caldwell, Kromhout and Schachter, 1981). In concordance *ompA* was the third most abundant transcript. In *C. trachomatis ompA* give rise to two transcripts that are differentially expressed during the life cycle (Stephens, Wagar and Edman, 1988). *C. pneumoniae* features even three TSSs for *ompA* located -266, -254 and -165 nts relative to the start codon, whereby only the -254 TSSs is conserved and also found in *C. trachomatis* (Albrecht *et al.*, 2011). The presence of these three TSSs in *C. pneumoniae ompA* found in the original dRNA-Seq study (Albrecht *et al.*, 2011) could be confirmed by the reanalysis. Interestingly both, *C. psittaci* and *C. abortus* harbor

tandem TSSs for *ompA* at -253 and -245 nts relative to the translational start site, indicating that the sTSS at position -253/-254 is conserved among chlamydial species. The *ompA* CDS is ~70% identical in *C*. *trachomatis*, *C*. *pneumoniae*, *C*. *psittaci* and *C*. *abortus* and the -254 TSS is conserved among all four species. These findings raised the question of the general conservation rate of ORFs and TSS positions (and consequently 5'UTR as well as promoter sequences) in *Chlamydia*.

The coding sequences of *Chlamydia* are similar, especially *C*. *psittaci* and *C*. *abortus* ORFs share ~95% sequence identity. Consequently, it has been indicated that the differing expression levels of transcripts, rather than sequence dissimilarities alone might contribute to the differences in host specificity and tissue tropism (Braukmann *et al.*, 2012). In concordance, the comparison of relative TSS positions revealed a high level of diversity. Varying TSS positions and consequently 5'UTR as well as promoter sequences (**Figure 14**) indicate different regulatory mechanisms which may ultimately result in the distinct expression observed for highly conserved genes. In summary the results show that among *Chlamydia*, positions of TSSs as well 5'UTR and promoter sequences are in general evolutionary less conserved than the corresponding ORF sequences (**Figure 14**). Overall, the results are in agreement with a comparison between *E*. *coli* and *Klebsiella pneumoniae*, which revealed that the ORF was most conserved, followed by the promoter and the 5'UTR region as the most diverse element (Kim *et al.*, 2012).

Knowing the precise TSS positions a comprehensive promoter motif analysis for *Chlamydia* could be performed. In total 2,157 promoters (60.57%) containing conserved motifs were found most of which are potential σ factor binding sites (**Figure 15**). The most often found motif (1,590) corresponds to σ^{66} whereas potential σ^{54} and σ^{28} binding sites occur less frequently. Furthermore, novel chlamydial motifs were found too (**Figure 16**) and a comparison of the four motifs to known ones using Tomtom (Gupta *et al.*, 2007) yielded no significant hits. This indicates that the found motifs have not been described yet and might be specific for phylum *Chlamydiae*. Moreover, the motifs are

conserved in *Chlamydia* since they were also found in *W. chondrophila*, a distant relative that diverged about one billion years ago (de Barsy and Greub, 2013).

In summary a workflow for reliable annotation of chlamydial TSSs as well as a comparative analysis of 5'UTRs and promoter motifs was presented. The results of this section were integrated in a manuscript that is drafted for publication.

6.4 Virulence factors are differentially expressed in chlamydial species

The close relatives *C. psittaci* and *C. abortus* share the vast majority of genes (916) and only few species-specific ORFs were found (**Figure 17**). Sequence comparison of the 54 unique genes in *C. psittaci* showed that 25 encode Pmps, predicted Incs or T3SE and are therefore putatively involved in virulence. The remaining 29 genes encode 24 hypothetical proteins or are involved in biosynthetic processes. Three of the genes unique for *C. psittaci* are encoded on the PZ, which is a genomic location near the replication terminus and a major source of diversity among chlamydial genomes (Read *et al.*, 2000). It has been shown the *C. psittaci* PZ spans about 29 kb and encodes 16 genes (Voigt, Schöfl and Saluz, 2012). It has less gene content than the respective PZs of *C. caviae* GPIC and *C. felis* FeC/-56 with 22 and 29 genes, respectively. However, it is larger than the PZ of *C. abortus* S26/3 that encodes 11 genes (Voigt, Schöfl and Saluz, 2012). This illustrates the diversity of the PZ among closely related chlamydial species. Moreover, two *C. psittaci* genes encoded on the PZ i.e., CPS0B_RS02845 and CPS0B_RS02850 are predicted to be T3SE and highly up-regulated in EBs (CPS0B_RS02850 is also enriched upon TEX treatment). This makes them very interesting candidates for novel VFs specific for *C. psittaci*.

Another difference between *C. psittaci*, *C. abortus* and *W. chondrophila* is the presence of a ~7.5 kb plasmid, which is conserved but present only in some *Chlamydia* species like *C. trachomatis* and *C.* muridarum. The presence of a plasmid in *C. psittaci* 02DC15 could be demonstrated by *de novo* assembly of unaligned reads and its sequence is identical to

plasmids of other *C*. *psittaci* strains. On this extrachromosomal DNA eight genes are encoded. Plasmid reads represented on average 0.85% of libraries (**Figure 7**) and three genes, *pgp1*, *pgp2* and *pgp8* are differentially expressed in *C*. *psittaci* EBs and RBs. Of these, the DNA helicases *pgp1* and pgp2 were more abundant in EBs, whereas the integrase *pgp8* was up-regulated in RBs. Up-regulation of the helicases might indicate regulatory functions of the plasmid on the transcription in *C*. *psittaci* EBs. Similar mechanisms were observed in *C*. *trachomatis* and *C*. *muridarum* in which plasmid genes are involved in the infectivity and virulence by regulating transcription of numerous genomic ORFs (Carlson *et al.*, 2008; O'Connell *et al.*, 2011). However, a study of the *C*. *psittaci* 6BC strain showed no effect of plasmid deletion in the infection of mice (Miyairi *et al.*, 2011).

Of the 16 *C*. *abortus* genes, which are not shared with *C*. *psittaci* and *W*. *chondrophila* (**Figure 17**) putatively VFs are the highly expressed Inc (CAB_RS03935) and a Pmp (CAB_RS01425). The remaining genes encode an adenylate kinase (CAB_RS03700) and 13 hypothetical proteins. Two of the hypothetical proteins i.e. CAB_RS01525 and CAB_RS01705 might be involved in infection because they are highly up-regulated in EBs.

W. *chondrophila* contains in total 1271 protein coding genes which are not shared with *C*. *psittaci* or *C*. *abortus* (**Figure 17**). Known *W*. *chondrophila* specific VFs participate in stress response, drug resistance or belong to the *W*. *chondrophila ompA* family. The 11 Omps encode putative porins, which might facilitate *W*. *chondrophila* infection (Bertelli *et al.*, 2010). However, their detailed functions are unknown. Interestingly, these *ompA* genes are partially shared with the *Chlamydiaceae* (Bertelli *et al.*, 2010) and for two, i.e. OmpA2 and OmpA3, adhesive properties on epithelial cells could be already demonstrated (Kebbi-Beghdadi *et al.*, 2015). In concordance *ompA2* and *ompA3* expression was up-regulated in *W*. *chondrophila* EBs and in general the majority of the Omps are differentially expressed in both developmental states.

Despite the presence of species-specific VFs the differential expression of homologous genes might have a great impact on host adaption, tissue tropism and virulence phenotype (Braukmann *et al.*, 2012). The inter-species comparison of gene expression revealed that among the 916 genes shared among *C. psittaci* and *C. abortus* (**Figure 17**), the 354 shared not with *W. chondrophila* correlate more in their expression (Pearson = 0.67) than the 562 core genes (Pearson = 0.49). This implies more distinct gene expression of core genes in *C. psittaci* and *C. abortus* that are mainly involved in essential cellular processes (**Supplementary Table 4**). Moreover, the core gene expression was more similar (Pearson = 0.75) among *C. abortus* and *W. chondrophila,* which might be due to similar metabolic requirements because of their ability to colonize female reproductive system and association with abortions (Dilbeck *et al.*, 1990; Henning *et al.*, 2002; Essig and Longbottom, 2015). Despite these differences also a core transcriptome in EBs and RBs was found that consisted of 209 (37%) and 327 (58%) similarly expressed genes in all three pathogens, respectively.

In addition to this general analysis, the expression of 52 known VFs and immunogenic genes shared between *C. psittaci, C. abortus* and *W. chondrophila* was compared (**Figure 18**, **Figure 19**). Two examples of VFs that are shared among *C. psittaci* and *C. abortus* but not with *W. chondrophila* are *sinC* and *tarP*. Both transcripts are higher expressed in *C. psittaci* compared to *C. abortus* (**Figure 18**), which might relate to some differences of the close relatives. TARP is translocated into the host cytosol upon EB attachment (Clifton *et al.*, 2004) and is the most abundant T3SE in *C. trachomatis* EBs (Saka *et al.*, 2011). Although the detailed mechanism of chlamydial entry has yet to be elucidated, evidence suggests that TARP mediates *Chlamydia* internalization (Betts, Wolf and Fields, 2009). SINC is conserved in *C. psittaci*, *C. abortus*, *C. caviae* and *C. felis*, but not in the major human pathogens *C. trachomatis* and *C. pneumoniae* (Mojica *et al.*, 2015). The factor is secreted and targets a conserved part of the inner nuclear membrane and also uninfected neighbors (Mojica *et al.*, 2015). On the contrary, *incA* and *incB* were higher expressed in *C. abortus* (**Figure 18**). Incs are important VFs as they are

located at the direct interphase between the inclusion and the host cytosol and inhibit fusion with endosomal compartments but promote fusogenicity with host vesicles (Li *et al.*, 2008). *C. psittaci* IncB associates with the host protein Snapin and connects chlamydial inclusions with the microtubule network (Böcker *et al.*, 2014). *C. psittaci* IncA interacts with the host protein G3BP1, leads to a decreased concentration of the c-Myc protein, and up-regulation of *incA* might contribute to increased inhibition of apoptosis and the prolonged life cycle of *C. abortus*. Other transcripts for VFs higher expressed in *C. abortus* are *pmpD* and *pmpH* (**Figure 18**) involved in chlamydial attachment and internalization (Wheelhouse *et al.*, 2009).

Among the 562 core genes that are shared among all three pathogens (**Figure 17**) 22 VFs and immunogenic genes were found of which most are differentially (e.g. *hctA* and *omcB*) and only very few were similarly expressed (e.g. *hrtA*) in EBs and RBs of all species (**Figure 19**). HctA is involved in chromosome condensation at the end of the developmental cycle (Barry, Hayes and Hackstadt, 1992) and high expression of *hctA* might accelerate condensation of DNA and thus EB formation. A higher expression of *hctA* was detected in EBs of *W. chondrophila*, which might contribute to the fast developmental cycle in HEp-2 cells (30h). OmcB mediates initial contact with the host cell (Fechtner *et al.*, 2013) and the transcript is enriched in *C. trachomatis* EBs (Albrecht *et al.*, 2010). Therefore, higher *omcB* expression in *C. psittaci* and *W. chondrophila* EBs (**Figure 19**) might directly influence attachment to the host cell and consequently, cause increased infectivity.

It is important to note that harvesting time points have to be considered when focusing on an inter-species comparison of RB and EB transcriptomes. RBs and EBs of the three pathogens were purified at different time points in order to isolate only characteristic developmental forms, e.g. by exclusion of RB to EB transition states, also referred to as intermediate bodies. Gene expression is a highly dynamic process and thus the expression results found in this work may not apply to forms at other time points during infection. However, time series transcriptomic analyses of the infectious and non-

infectious states are possible due to the presented protocol for EB and RB purification and ever decreasing sequencing costs. The presented differential gene expression analysis with a focus on virulence related genes was published as part of the presented work (Beder and Saluz, 2018).

6.5 Conclusion and prospects

This work demonstrated a possible workflow and allowed first insights into the differential gene expression within *Chlamydia*. It was shown that the replicative and infectious chlamydial states express distinctive transcriptomes and the cellular processes emphasized differ. Moreover, the very first chlamydial inter-species transcriptome comparison was presented, which revealed similarities but also considerable differences in the expression of homologous genes. This demonstrates that despite the resemblances on the genome level, genes are differentially expressed by the close relatives *C. psittaci* and *C. abortus* and this might influence infectivity, host specificity and tissue tropism of the pathogens. A promising next step would be the analysis of VF expression on the protein level and to study potential new VFs highlighted in this work.

The distinct expression observed for homologous genes might be caused by different gene regulatory mechanisms and therefore a detailed annotation of chlamydial TSSs, including 5’UTR and promoter sequences is of great importance. Reliable prediction of chlamydial TSSs was achieved by optimizing and intersecting results of three TSS annotation tools. Knowing the TSS locations it was possible to analyze their relative positions to genes and also their conservation across species. Interestingly, TSS positions and consequently 5'UTRs of individual transcripts are evolutionary less conserved than the corresponding ORF sequences. Indeed, these differences indicate that distinct gene regulatory mechanisms are responsible for the differential expression levels of conserved genes. In the past decade various targeted transformation methods and mutant libraries for *Chlamydia* have been established and with technologies like the CRISPR/Cas further experiments might include the exchange of regulatory elements like

the 5'UTR or promoter sequences within and among chlamydial species to study the influence on the transcript expression.

Overall, the presented protocol for EB and RB purification is straightforward and guaranteed integrity of the *Chlamydia*. Especially, in combination with the removal of rRNAs and decreasing costs for RNA-Seq it has a high potential to be used for more advanced experiments e.g. time series analyses or drug treatments. Moreover, proteome experiments might be carried out to analyze the gene expression on the protein level and also to test the overall similarity to the transcriptome. Another possibility might be the application of Ribo-Zero™ to RNA directly isolated from infected cells. Ribo-Zero™ was shown in this work to efficiently remove chlamydial rRNAs and an immediate lysis of the cells would ensure an unaltered transcriptome due to less experimental handling of the sample.

Other very interesting experiments might include the analysis of more chlamydial species like *C. trachomatis* and C. *pneumoniae*. This was initially planned but since RNA-Seq drastically improved over the past years the available data did not match the current standard for differential gene expression analysis. However, new dRNA-Seq data could be generated. Also, very interesting would be the transcriptome analysis of closely related chlamydial isolates with distinct niche adaptions. For several species like e.g. *C. trachomatis*, *C. pneumoniae*, *C. psittaci* and *C. suis* genomes from various isolates are sequenced and an analysis of closely related species would allow the correlation of genomic polymorphisms to the transcriptomes. Eventually, it might be possible to capture transcriptional dynamics and relate them specifically to the distinct host adaptations and virulence phenotypes.

In summary, this work gave valuable insights in the transcriptome expression of infectious and non-infectious chlamydial states and indicated regulatory mechanisms responsible for these.

7 References

Abdelrahman, Y. M. and Belland, R. J. (2005) 'The chlamydial developmental cycle.', *FEMS microbiology reviews*, 29(5), pp. 949–59. doi: 10.1016/j.femsre.2005.03.002.

Agarwal, A. *et al.* (2010) 'Comparison and calibration of transcriptome data from RNA-Seq and tiling arrays.', *BMC genomics*, 11, p. 383. doi: 10.1186/1471-2164-11-383.

Albrecht, M. *et al.* (2010) 'Deep sequencing-based discovery of the Chlamydia trachomatis transcriptome.', *Nucleic acids research*, 38(3), pp. 868–77. doi: 10.1093/nar/gkp1032.

Albrecht, M. *et al.* (2011) 'The transcriptional landscape of Chlamydia pneumoniae.', *Genome biology*. BioMed Central Ltd, 12(10), p. R98. doi: 10.1186/gb-2011-12-10-r98.

Amann, R. *et al.* (1997) 'Obligate intracellular bacterial parasites of acanthamoebae related to Chlamydia spp.', *Applied and environmental microbiology*, 63(1), pp. 115–21.

Amman, F. *et al.* (2014) 'TSSAR: TSS annotation regime for dRNA-seq data.', *BMC bioinformatics*, 15(1), p. 89. doi: 10.1186/1471-2105-15-89.

Arnold, R. *et al.* (2009) 'Sequence-Based Prediction of Type III Secreted Proteins', *PLoS Pathogens*. Edited by C. E. Stebbins, 5(4), p. e1000376. doi: 10.1371/journal.ppat.1000376.

Van Assche, E. *et al.* (2015) 'RNA-binding proteins involved in post-transcriptional regulation in bacteria', *Frontiers in Microbiology*, 6. doi: 10.3389/fmicb.2015.00141.

Aziz, R. K. *et al.* (2008) 'The RAST Server: rapid annotations using subsystems technology.', *BMC genomics*, 9(1), p. 75. doi: 10.1186/1471-2164-9-75.

Bailey, T. L. *et al.* (2015) 'The MEME Suite', *Nucleic Acids Research*, 43(W1), pp. W39–W49. doi: 10.1093/nar/gkv416.

Balin, B. J. *et al.* (1998) 'Identification and localization of Chlamydia pneumoniae in the Alzheimer's brain.', *Medical microbiology and immunology*, 187(1), pp. 23–42.

Barkallah, M. *et al.* (2013) 'First detection of Waddlia chondrophila in Africa using SYBR Green real-time PCR on veterinary samples', *Veterinary Microbiology*. Elsevier B.V., 164(1–2), pp. 101–107. doi: 10.1016/j.vetmic.2013.01.036.

Barrios, H., Valderrama, B. and Morett, E. (1999) 'Compilation and analysis of sigma(54)-dependent promoter sequences.', *Nucleic acids research*, 27(22), pp. 4305–13.

Barry, C. E., Hayes, S. F. and Hackstadt, T. (1992) 'Nucleoid condensation in Escherichia coli that express a chlamydial histone homolog.', *Science (New York, N.Y.)*, 256(5055), pp. 377–9. doi: 10.1126/science.256.5055.377.

de Barsy, M. and Greub, G. (2013) 'Waddlia chondrophila: from biology to pathogenicity.', *Microbes and infection*. Elsevier Masson SAS, 15(14–15), pp. 1033–41. doi: 10.1016/j.micinf.2013.09.010.

Bastidas, R. J. *et al.* (2013) 'Chlamydial intracellular survival strategies.', *Cold Spring Harbor perspectives in medicine*, 3(5), p. a010256. doi: 10.1101/cshperspect.a010256.

Baud, D. *et al.* (2011) 'Role of Chlamydia trachomatis in miscarriage.', *Emerging infectious diseases*, 17(9), pp. 1630–5. doi: 10.3201/eid1709.100865.

Baud, D. *et al.* (2011) 'Waddlia chondrophila: From Bovine Abortion to Human Miscarriage', *Clinical Infectious Diseases*, 52(12), pp. 1469–1471. doi: 10.1093/cid/cir205.

Baud, D., Osterheld, M., *et al.* (2014) 'Role of Waddlia chondrophila Placental Infection in Miscarriage', 20(3).

Baud, D., Goy, G., *et al.* (2014) 'Role of Waddlia chondrophila Placental Infection in Miscarriage', *Emerging Infectious Diseases*, 20(3), pp. 460–464. doi: 10.3201/eid2003.131019.

Baud, D., Greub, G. and Centre, H. (2011) 'Intracellular bacteria and adverse pregnancy outcomes'.

Beder, T. *et al.* (2016) 'Purification of infectious and non-infectious chlamydial particles using iodixanol for density gradient preparation.', *Journal of microbiological methods*, 128(3), pp. 20–23. doi: 10.1016/j.mimet.2016.06.029.

Beder, T. and Saluz, H. P. (2018) 'Virulence-related comparative transcriptomics of infectious and non-infectious chlamydial particles', *BMC Genomics*. BMC Genomics, 19(1), p. 575. doi: 10.1186/s12864-018-4961-x.

Bedson, S. P. and Bland, J. O. W. (1932) 'A morphological study of psittacosis virus, with the description of a developmental cycle.', *Br. J. Exp. Pathol.*, 13, pp. 461–466.

Bedson, S. P., Western, G. T. and Simpson, S. L. (1930) 'Observations on the aetiology of psittacosis.', *Lancet*, i, pp. 235–236.

Beeckman, D. S. A. and Vanrompay, D. C. G. (2009) 'Zoonotic Chlamydophila psittaci infections from a clinical perspective', *Clinical Microbiology and Infection*. European Society of Clinical Infectious Diseases, 15(1), pp. 11–17. doi: 10.1111/j.1469-0691.2008.02669.x.

Belland, R. J., Zhong, G., *et al.* (2003) 'Genomic transcriptional profiling of the developmental cycle of Chlamydia trachomatis', *Proceedings of the National Academy of Sciences*, 100(14), pp. 8478–8483. doi: 10.1073/pnas.1331135100.

Belland, R. J., Nelson, D. E., *et al.* (2003) 'Transcriptome analysis of chlamydial growth during IFN-gamma-mediated persistence and reactivation.', *Proceedings of the National Academy of Sciences of the United States of America*, 100(26), pp. 15971–6. doi: 10.1073/pnas.2535394100.

Benjamini, Y. *et al.* (2001) 'Controlling the false discovery rate in behavior genetics research.', *Behavioural brain research*, 125(1–2), pp. 279–84. doi: 10.2307/2346101.

Bertelli, C. *et al.* (2010) 'The waddlia genome: A window into chlamydial biology', *PLoS ONE*, 5(5). doi: 10.1371/journal.pone.0010890.

Betts-Hampikian, H. J. and Fields, K. A. (2010) 'The Chlamydial Type III Secretion Mechanism: Revealing Cracks in a Tough Nut.', *Frontiers in microbiology*, 1, p. 114. doi: 10.3389/fmicb.2010.00114.

Betts, H. J., Wolf, K. and Fields, K. a. (2009) 'Effector protein modulation of host cells: examples in the Chlamydia spp. arsenal', *Current Opinion in Microbiology*, 12(1), pp. 81–87. doi: 10.1016/j.mib.2008.11.009.

Binet, R. and Maurelli, A. T. (2005) 'Fitness cost due to mutations in the 16S rRNA associated with spectinomycin resistance in Chlamydia psittaci 6BC.', *Antimicrobial agents and chemotherapy*, 49(11), pp. 4455–64. doi: 10.1128/AAC.49.11.4455-4464.2005.

Binet, R. and Maurelli, A. T. (2009) 'Transformation and isolation of allelic exchange mutants of Chlamydia psittaci using recombinant DNA introduced by electroporation', *Proceedings of the National Academy of Sciences*, 106(1), pp. 292–297. doi: 10.1073/pnas.0806768106.

Bischler, T. *et al.* (2015) 'Differential RNA-seq (dRNA-seq) for annotation of transcriptional start sites and small RNAs in Helicobacter pylori.', *Methods (San Diego, Calif.)*, 86, pp. 89–101. doi: 10.1016/j.ymeth.2015.06.012.

Blasi, F. *et al.* (1998) 'Epidemiology of Chlamydia pneumoniae.', *Clinical microbiology and infection : the official publication of the European Society of Clinical Microbiology and Infectious Diseases*, 4 Suppl 4, pp. S1–S6.

Böcker, S. *et al.* (2014) 'Chlamydia psittaci inclusion membrane protein IncB associates with host protein Snapin.', *International journal of medical microbiology : IJMM*, 304(5–6), pp. 542–53. doi: 10.1016/j.ijmm.2014.03.005.

Bolger, A. M., Lohse, M. and Usadel, B. (2014) 'Trimmomatic: a flexible trimmer for Illumina sequence data.', *Bioinformatics (Oxford, England)*, 30(15), pp. 2114–20. doi: 10.1093/bioinformatics/btu170.

Bonocora, R. P. *et al.* (2015) 'Genome-Scale Mapping of Escherichia coli σ54 Reveals Widespread, Conserved Intragenic Binding.', *PLoS genetics*, 11(10), p. e1005552. doi: 10.1371/journal.pgen.1005552.

Borel, N. *et al.* (2014) 'The role of viable but non-infectious developmental forms in chlamydial biology.', *Frontiers in cellular and infection microbiology*, 4, p. 97. doi: 10.3389/fcimb.2014.00097.

Borel, N., Polkinghorne, A. and Pospischil, A. (2018) 'A Review on Chlamydial Diseases in Animals: Still a Challenge for Pathologists?', *Veterinary Pathology*, 55(3), pp. 374–390. doi: 10.1177/0300985817751218.

Bose, S. K. and Paul, R. G. (1982) 'Purification of Chlamydia trachomatis lymphogranuloma venereum elementary bodies and their interaction with HeLa cells.', *Journal of general microbiology*, 128(6), pp. 1371–9. doi: 10.1099/00221287-128-6-1371.

Bovarnick, M. R., Miller, J. C. and Snyder, J. C. (1950) 'The influence of certain salts, amino acids, sugars, and proteins on the stability of rickettsiae.', *Journal of bacteriology*, 59(4), pp. 509–22.

Braukmann, M. *et al.* (2012) 'Distinct intensity of host-pathogen interactions in Chlamydia psittaci- and Chlamydia abortus-infected chicken embryos', *Infection and Immunity*. Edited by J. B. Bliska, 80(9), pp. 2976–2988. doi: 10.1128/IAI.00437-12.

Buttner, D. (2012) 'Protein Export According to Schedule: Architecture, Assembly, and Regulation of Type III Secretion Systems from Plant- and Animal-Pathogenic Bacteria', *Microbiology and Molecular Biology Reviews*, 76(2), pp. 262–310. doi: 10.1128/MMBR.05017-11.

Byrne, G. I. (2010) 'Chlamydia trachomatis strains and virulence: rethinking links to infection prevalence and disease severity.', *The Journal of infectious diseases*, 201 Suppl, pp. S126-33. doi: 10.1086/652398.

Caldwell, H. D., Kromhout, J. and Schachter, J. (1981) 'Purification and partial characterization of the major outer membrane protein of Chlamydia trachomatis.', *Infection and immunity*, 31(3), pp. 1161–76. doi: PMC.

Carabeo, R. A., Mead, D. J. and Hackstadt, T. (2003) 'Golgi-dependent transport of cholesterol to the Chlamydia trachomatis inclusion.', *Proceedings of the National Academy of Sciences of the United States of America*, 100(11), pp. 6771–6. doi: 10.1073/pnas.1131289100.

Carlson, J. H. *et al.* (2008) 'The Chlamydia trachomatis Plasmid Is a Transcriptional Regulator of Chromosomal Genes and a Virulence Factor', *Infection and Immunity*, 76(6), pp. 2273–2283. doi: 10.1128/IAI.00102-08.

Celesnik, H., Deana, A. and Belasco, J. G. (2007) 'Initiation of RNA decay in Escherichia coli by 5' pyrophosphate removal.', *Molecular cell*, 27(1), pp. 79–90. doi: 10.1016/j.molcel.2007.05.038.

Charif, D. and Lobry, J. R. (2007) 'SeqinR 1.0-2: A Contributed Package to the R Project for Statistical Computing Devoted to Biological Sequences Retrieval and Analysis', in Bastolla, U. et al. (eds) *Structural Approaches to Sequence Evolution: Molecules, Networks, Populations*. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 207–232. doi: 10.1007/978-3-540-35306-5_10.

Chen, Q. *et al.* (2014) 'Seroprevalence of Chlamydophila abortus infection in yaks (Bos grunniens) in Qinghai, China.', *Tropical animal health and production*, 46(3), pp. 503–7. doi: 10.1007/s11250-013-0519-8.

Clifton, D. R. *et al.* (2004) 'From The Cover: A chlamydial type III translocated protein is tyrosine-phosphorylated at the site of entry and associated with recruitment of actin', *Proceedings of the National Academy of Sciences*, 101(27), pp. 10166–10171. doi: 10.1073/pnas.0402829101.

Coles, A. C. (1930) 'Microorganisms in psittacosis.', *Lancet*, i, pp. 1011–1012.

Collingro, A. *et al.* (2005) 'Recovery of an environmental Chlamydia strain from activated sludge by co-cultivation with Acanthamoeba sp.', *Microbiology (Reading, England)*, 151(Pt 1), pp. 301–9. doi: 10.1099/mic.0.27406-0.

Collingro, A. *et al.* (2011) 'Unity in variety--the pan-genome of the Chlamydiae.', *Molecular biology and evolution*, 28(12), pp. 3253–70. doi: 10.1093/molbev/msr161.

Crooks, G. E. *et al.* (2004) ‘WebLogo: a sequence logo generator.’, *Genome research*, 14(6), pp. 1188–90. doi: 10.1101/gr.849004.

Crosse, B. A. (1990) ‘Psittacosis: a clinical review.’, *The Journal of infection*, 21(3), pp. 251–9. doi: http://doi.org/10.1016/0163-4453(90)93909-C.

Dean, D., Meyers, G. S. and Read, T. D. (2006) ‘Lessons and challenges arising from the “first wave” of Chlamydia genome sequencing. In: Chlamydia: Genomics and Pathogenesis, (Ed. by P. M. Bavoil & P.B. Wyrick)’, *Norfolk, UK: Horizon Bioscience*, pp. 1–24.

Dilbeck-Robertson, P. *et al.* (2003) ‘Results of a new serologic test suggest an association of Waddlia chondrophila with bovine abortion.’, *Journal of veterinary diagnostic investigation : official publication of the American Association of Veterinary Laboratory Diagnosticians, Inc*, 15(6), pp. 568–569. doi: 10.1177/104063870301500609.

Dilbeck, P. M. *et al.* (1990) ‘Isolation of a previously undescribed rickettsia from an aborted bovine fetus’, *Journal of Clinical Microbiology*, 28(4), pp. 814–816.

Dugar, G. *et al.* (2013) ‘High-Resolution Transcriptome Maps Reveal Strain-Specific Regulatory Features of Multiple Campylobacter jejuni Isolates’, *PLoS Genetics*. Edited by D. Hughes, 9(5), p. e1003495. doi: 10.1371/journal.pgen.1003495.

Durand, N. J., Nicolas, J. and Favre, M. (1913) ‘Lymphogranulomatose inguinale subaiguë d’origine génitale probable, peut-être vénérienne.’, *Bulletin de la Société des Médecins des Hôpitaux de Paris*, 35, pp. 274–288.

Eckweiler, D. *et al.* (2018) ‘PRODORIC2: the bacterial gene regulation database in 2018.’, *Nucleic acids research*, 46(D1), pp. D320–D326. doi: 10.1093/nar/gkx1091.

Elwell, C., Mirrashidi, K. and Engel, J. (2016) ‘Chlamydia cell biology and pathogenesis’, *Nature Reviews Microbiology*, 14(6), pp. 385–400. doi: 10.1038/nrmicro.2016.30.

Engstrom, P. *et al.* (2013) ‘Mutations in hemG Mediate Resistance to Salicylidene Acylhydrazides, Demonstrating a Novel Link between Protoporphyrinogen Oxidase (HemG) and Chlamydia trachomatis Infectivity’, *Journal of Bacteriology*, 195(18), pp. 4221–4230. doi: 10.1128/JB.00506-13.

Essig, A. and Longbottom, D. (2015) ‘Chlamydia abortus: New Aspects of Infectious Abortion in Sheep and Potential Risk for Pregnant Women’, *Current Clinical Microbiology Reports*, 2(1), pp. 22–34. doi: 10.1007/s40588-015-0014-2.

Ettwiller, L. *et al.* (2016) ‘A novel enrichment strategy reveals unprecedented number of novel transcription start sites at single base resolution in a model prokaryote and the gut microbiome.’, *BMC genomics*, 17, p. 199. doi: 10.1186/s12864-016-2539-z.

Ewels, P. *et al.* (2016) ‘MultiQC: summarize analysis results for multiple tools and samples in a single report.’, *Bioinformatics (Oxford, England)*, 32(19), pp. 3047–8. doi: 10.1093/bioinformatics/btw354.

Fahr, M. J. *et al.* (1995) ‘Characterization of late gene promoters of Chlamydia trachomatis.’, *Journal of bacteriology*, 177(15), pp. 4252–60.

Fechtner, T. *et al.* (2013) 'Characterization of the interaction between the chlamydial adhesin omcb and the human host cell', *Journal of Bacteriology*, 195(23), pp. 5323–5333. doi: 10.1128/JB.00780-13.

Fields, K. A. and Hackstadt, T. (2002) 'The chlamydial inclusion: escape from the endocytic pathway.', *Annual review of cell and developmental biology*, 18(1), pp. 221–45. doi: 10.1146/annurev.cellbio.18.012502.105845.

Flaxman, S. R. *et al.* (2017) 'Global causes of blindness and distance vision impairment 1990–2020: a systematic review and meta-analysis', *The Lancet Global Health*, 5(12), pp. e1221–e1234. doi: 10.1016/S2214-109X(17)30393-5.

Friis, R. R. (1972) 'Interaction of L cells and Chlamydia psittaci: entry of the parasite and host responses to its development.', *Journal of bacteriology*, 110(2), pp. 706–21.

Georg, J. and Hess, W. R. (2011) 'cis-antisense RNA, another level of gene regulation in bacteria.', *Microbiology and molecular biology reviews : MMBR*, 75(2), pp. 286–300. doi: 10.1128/MMBR.00032-10.

Giannoukos, G. *et al.* (2012) 'Efficient and robust RNA-seq process for cultured bacteria and complex community transcriptomes.', *Genome biology*, 13(3), p. R23. doi: 10.1186/gb-2012-13-3-r23.

Giménez, D. F. (1964) 'STAINING RICKETTSIAE IN YOLK-SAC CULTURES.', *Stain technology*, 39, pp. 135–40.

Global Burden of Disease Study 2013 Collaborators (2015) 'Global, regional, and national incidence, prevalence, and years lived with disability for 301 acute and chronic diseases and injuries in 188 countries, 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013.', *Lancet (London, England)*, 386(9995), pp. 743–800. doi: 10.1016/S0140-6736(15)60692-4.

Goodwin, S., McPherson, J. D. and McCombie, W. R. (2016) 'Coming of age: ten years of next-generation sequencing technologies', *Nature Reviews Genetics*, 17(6), pp. 333–351. doi: 10.1038/nrg.2016.49.

Goy, G. and Greub, G. (2009) 'Antibiotic susceptibility of Waddlia chondrophila in Acanthamoeba castellanii amoebae', *Antimicrobial Agents and Chemotherapy*, 53(6), pp. 2663–2666. doi: 10.1128/AAC.00046-09.

Grabherr, M. G. *et al.* (2011) 'Full-length transcriptome assembly from RNA-Seq data without a reference genome', *Nature Biotechnology*, 29(7), pp. 644–652. doi: 10.1038/nbt.1883.

Graham, J., Ford, T. and Rickwood, D. (1994) 'The preparation of subcellular organelles from mouse liver in self-generated gradients of iodixanol.', *Analytical biochemistry*, 220(2), pp. 367–73. doi: 10.1006/abio.1994.1351.

Grant, C. E., Bailey, T. L. and Noble, W. S. (2011) 'FIMO: scanning for occurrences of a given motif.', *Bioinformatics (Oxford, England)*, 27(7), pp. 1017–8. doi: 10.1093/bioinformatics/btr064.

de Groot, A. *et al.* (2014) ‘RNA sequencing and proteogenomics reveal the importance of leaderless mRNAs in the radiation-tolerant bacterium Deinococcus deserti.’, *Genome biology and evolution*, 6(4), pp. 932–48. doi: 10.1093/gbe/evu069.

Gupta, S. *et al.* (2007) ‘Quantifying similarity between motifs.’, *Genome biology*, 8(2), p. R24. doi: 10.1186/gb-2007-8-2-r24.

Haas, B. J. *et al.* (2012) ‘How deep is deep enough for RNA-Seq profiling of bacterial transcriptomes?’, *BMC Genomics*, 13(1), p. 734. doi: 10.1186/1471-2164-13-734.

Hackstadt, T. *et al.* (1996) ‘Chlamydia trachomatis interrupts an exocytic pathway to acquire endogenously synthesized sphingomyelin in transit from the Golgi apparatus to the plasma membrane.’, *The EMBO journal*, 15(5), pp. 964–77.

Hadfield, J. *et al.* (2017) ‘Comprehensive global genome dynamics of Chlamydia trachomatis show ancient diversification followed by contemporary mixing and recent lineage expansion’, *Genome Research*, 27(7), pp. 1220–1229. doi: 10.1101/gr.212647.116.

Hahn, D. L., Dodge, R. W. and Golubjatnikov, R. (1991) ‘Association of Chlamydia pneumoniae (strain TWAR) infection with wheezing, asthmatic bronchitis, and adult-onset asthma.’, *JAMA*, 266(2), pp. 225–30.

Haider, S. *et al.* (2008) ‘Chlamydia-like bacteria in respiratory samples of community-acquired pneumonia patients.’, *FEMS microbiology letters*, 281(2), pp. 198–202. doi: 10.1111/j.1574-6968.2008.01099.x.

Halberstaedter, L. and Prowazek, S. (1907) ‘Zur atiologie des trachoms.’, *Deutsche Medizinische Wochenschrift*, 33, pp. 1285–1287.

Harris, S. R. *et al.* (2012) ‘Whole-genome analysis of diverse Chlamydia trachomatis strains identifies phylogenetic relationships masked by current clinical typing.’, *Nature genetics*, 44(4), pp. 413–9, S1. doi: 10.1038/ng.2214.

Hartley, J. C. *et al.* (2001) ‘Conjunctivitis due to Chlamydophila felis (Chlamydia psittaci feline pneumonitis agent) acquired from a cat: case report with molecular characterization of isolates from the patient and cat.’, *The Journal of infection*, 43(1), pp. 7–11. doi: 10.1053/jinf.2001.0845.

Hatch, T. P., Miceli, M. and Sublett, J. E. (1986) ‘Synthesis of disulfide-bonded outer membrane proteins during the developmental cycle of Chlamydia psittaci and Chlamydia trachomatis.’, *Journal of bacteriology*, 165(2), pp. 379–85. doi: 10.1128/jb.165.2.379-385.1986.

Hatch, T. P. (1996) ‘Disulfide cross-linked envelope proteins: the functional equivalent of peptidoglycan in chlamydiae?’, *Journal of bacteriology*, 178(1), pp. 1–5.

Hatch, T. P. (1999) ‘Developmental biology. Chlamydia (Stephens RS, ed)’, *American Society for Microbiology, Washington, DC*, pp. 29–67.

Henning, K. *et al.* (2002) ‘Neospora caninum and Waddlia chondrophila strain 2032/99 in a septic stillborn calf’, *Veterinary Microbiology*, 85(3), pp. 285–292.

Hooppaw, A. J. and Fisher, D. J. (2016) 'A Coming of Age Story: Chlamydia in the Post-Genetic Era', *Infection and Immunity*. Edited by A. T. Maurelli, 84(3), pp. 612–621. doi: 10.1128/IAI.01186-15.

Horn, M. *et al.* (2004) 'Illuminating the evolutionary history of chlamydiae.', *Science (New York, N.Y.)*, 304(5671), pp. 728–30. doi: 10.1126/science.1096330.

Horn, M. (2008) 'Chlamydiae as Symbionts in Eukaryotes', *Annual Review of Microbiology*, 62(1), pp. 113–131. doi: 10.1146/annurev.micro.62.081307.162818.

Howard, L., Orenstein, N. S. and King, N. W. (1974) 'Purification on renografin density gradients of Chlamydia trachomatis grown in the yolk sac of eggs.', *Applied microbiology*, 27(1), pp. 102–106.

Humphrys, M. S. *et al.* (2013) 'Simultaneous transcriptional profiling of bacteria and their host cells.', *PloS one*, 8(12), p. e80597. doi: 10.1371/journal.pone.0080597.

Islam, M. M. *et al.* (2019) 'In vitro analysis of genetically distinct Chlamydia pecorum isolates reveals key growth differences in mammalian epithelial and immune cells', *Veterinary Microbiology*, 232, pp. 22–29. doi: 10.1016/j.vetmic.2019.03.024.

Jewett, T. J. *et al.* (2010) 'The conserved Tarp actin binding domain is important for chlamydial invasion.', *PLoS pathogens*, 6(7), p. e1000997. doi: 10.1371/journal.ppat.1000997.

Jorjani, H. and Zavolan, M. (2014) 'TSSer: an automated method to identify transcription start sites in prokaryotic genomes from differential RNA sequencing data.', *Bioinformatics (Oxford, England)*, 30(7), pp. 971–4. doi: 10.1093/bioinformatics/btt752.

Kahane, S., Metzer, E. and Friedman, M. G. (1995) 'Evidence that the novel microorganism "Z" may belong to a new genus in the family Chlamydiaceae.', *FEMS microbiology letters*, 126(2), pp. 203–7.

Kalman, S. *et al.* (1999) 'Comparative genomes of Chlamydia pneumoniae and C. trachomatis.', *Nature genetics*, 21(4), pp. 385–9. doi: 10.1038/7716.

Kebbi-Beghdadi, C. *et al.* (2015) 'OmpA family proteins and Pmp-like autotransporter: new adhesins of Waddlia chondrophila.', *Pathogens and disease*, 73(6), p. ftv035. doi: 10.1093/femspd/ftv035.

Kim, D. *et al.* (2012) 'Comparative Analysis of Regulatory Elements between Escherichia coli and Klebsiella pneumoniae by Genome-Wide Transcription Start Site Profiling', *PLoS Genetics*. Edited by D. Hughes, 8(8), p. e1002867. doi: 10.1371/journal.pgen.1002867.

Knittler, M. R. *et al.* (2014) 'Chlamydia psittaci: new insights into genomic diversity, clinical pathology, host-pathogen interaction and anti-bacterial immunity.', *International journal of medical microbiology : IJMM*. Elsevier GmbH., 304(7), pp. 877–93. doi: 10.1016/j.ijmm.2014.06.010.

Kopf, M. *et al.* (2014) 'Comparative analysis of the primary transcriptome of Synechocystis sp. PCC 6803.', *DNA research : an international journal for rapid publication of reports on genes and genomes*, 21(5), pp. 527–39. doi: 10.1093/dnares/dsu018.

Kröger, C. *et al.* (2012) 'The transcriptional landscape and small RNAs of Salmonella enterica serovar Typhimurium.', *Proceedings of the National Academy of Sciences of the United States of America*, 109(20), pp. E1277-86. doi: 10.1073/pnas.1201061109.

Kuo, C. C. *et al.* (1993) 'Demonstration of Chlamydia pneumoniae in atherosclerotic lesions of coronary arteries.', *The Journal of infectious diseases*, 167(4), pp. 841–9.

Kuo, C. C. *et al.* (1995) 'Chlamydia pneumoniae (TWAR).', *Clinical microbiology reviews*, 8(4), pp. 451–61.

Lagae, S. *et al.* (2014) 'Emerging Chlamydia psittaci infections in chickens and examination of transmission to humans.', *Journal of medical microbiology*, 63(Pt 3), pp. 399–407. doi: 10.1099/jmm.0.064675-0.

Langmead, B. and Salzberg, S. L. (2012) 'Fast gapped-read alignment with Bowtie 2.', *Nature methods*, 9(4), pp. 357–9. doi: 10.1038/nmeth.1923.

Last, A. R. *et al.* (2018) 'Population-based analysis of ocular Chlamydia trachomatis in trachoma-endemic West African communities identifies genomic markers of disease severity', *Genome Medicine*, 10(1), p. 15. doi: 10.1186/s13073-018-0521-x.

Lawrence, M. *et al.* (2013) 'Software for Computing and Annotating Genomic Ranges', *PLoS Computational Biology*. Edited by A. Prlic, 9(8), p. e1003118. doi: 10.1371/journal.pcbi.1003118.

Li, D. *et al.* (2005) 'High-yield culture and purification of Chlamydiaceae bacteria.', *Journal of microbiological methods*, 61(1), pp. 17–24. doi: 10.1016/j.mimet.2004.10.020.

Li, Z. *et al.* (2008) 'Characterization of Fifty Putative Inclusion Membrane Proteins Encoded in the Chlamydia trachomatis Genome', *Infection and Immunity*, 76(6), pp. 2746–2757. doi: 10.1128/IAI.00010-08.

Lillie, R. D. (1930) 'Psittacosis: rickettsia-like inclusions in man and in experimental animals.', *Public Health Report*, 45, pp. 773–778.

Lin, Y. fei *et al.* (2013) 'A combination of improved differential and global RNA-seq reveals pervasive transcription initiation and events in all stages of the life-cycle of functional RNAs in Propionibacterium acnes, a major contributor to wide-spread human disease', *BMC Genomics*. BMC Genomics, 14(1), p. 620. doi: 10.1186/1471-2164-14-620.

Lindenbach, B. D. *et al.* (2005) 'Complete Replication of Hepatitis C Virus in Cell Culture', (July), pp. 623–627.

Liu, X. *et al.* (2010) 'Identification of Chlamydia trachomatis outer membrane complex proteins by differential proteomics.', *Journal of bacteriology*, 192(11), pp. 2852–60. doi: 10.1128/JB.01628-09.

Longbottom, D. and Coulter, L. J. (2003) 'Animal Chlamydioses and Zoonotic Implications', *Journal of Comparative Pathology*, 128(4), pp. 217–244. doi: 10.1053/jcpa.2002.0629.

Love, M. I., Huber, W. and Anders, S. (2014) 'Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2.', *Genome biology*, 15(12), p. 550. doi: 10.1186/s13059-014-0550-8.

Marques, P. X. *et al.* (2010) 'Identification of immunologically relevant proteins of Chlamydophila abortus using sera from experimentally infected pregnant ewes', *Clinical and Vaccine Immunology*, 17(8), pp. 1274–1281. doi: 10.1128/CVI.00163-10.

Mäurer, A. P. *et al.* (2007) 'Gene expression profiles of Chlamydophila pneumoniae during the developmental cycle and iron depletion-mediated persistence.', *PLoS pathogens*, 3(6), p. e83. doi: 10.1371/journal.ppat.0030083.

McClarty, G. (1994) 'Chlamydiae and the biochemistry of intracellular parasitism', *Trends in Microbiology*, 2(5), pp. 157–164. doi: 10.1016/0966-842X(94)90665-3.

McClenaghan, M., Herring, A. J. and Aitken, I. D. (1984) 'Comparison of Chlamydia psittaci isolates by DNA restriction endonuclease analysis.', *Infection and immunity*, 45(2), pp. 384–9.

Metzker, M. L. (2010) 'Sequencing technologies - the next generation.', *Nature reviews. Genetics*. Nature Publishing Group, 11(1), pp. 31–46. doi: 10.1038/nrg2626.

Miyagawa, Y. *et al.* (1935) 'Fourth report: studies on the virus of lymphogranuloma inguinale Nicolas, Favre and Durand. Cultivation of the virus on the chrorioallantoic membrane of the chicken embryo.', *Jpn. J. Exp. Med.*, 13, pp. 733–738.

Miyairi, I. *et al.* (2011) 'Chlamydia psittaci genetic variants differ in virulence by modulation of host immunity.', *The Journal of infectious diseases*, 204(4), pp. 654–63. doi: 10.1093/infdis/jir333.

Mojica, S. A. *et al.* (2015) 'SINC, a type III secreted protein of Chlamydia psittaci, targets the inner nuclear membrane of infected cells and uninfected neighbors', *Molecular Biology of the Cell*, 26(10), pp. 1918–1934. doi: 10.1091/mbc.E14-11-1530.

Moulder, J. W. (1966) 'The relation of the psittacosis group (Chlamydiae) to bacteria and viruses.', *Annual review of microbiology*, 20, pp. 107–30. doi: 10.1146/annurev.mi.20.100166.000543.

Mueller, K. E., Plano, G. V. and Fields, K. A. (2014) 'New frontiers in type III secretion biology: the Chlamydia perspective.', *Infection and immunity*, 82(1), pp. 2–9. doi: 10.1128/IAI.00917-13.

Mukhopadhyay, S. *et al.* (2004) 'Detailed protocol for purification of Chlamydia pneumoniae elementary bodies', *Journal of Clinical Microbiology*, 42(7), pp. 3288–3290. doi: 10.1128/JCM.42.7.3288.

Newman, L. *et al.* (2015) 'Global Estimates of the Prevalence and Incidence of Four Curable Sexually Transmitted Infections in 2012 Based on Systematic Review and Global Reporting.', *PloS one*, 10(12), p. e0143304. doi: 10.1371/journal.pone.0143304.

Nguyen, B. D. *et al.* (2011) 'Lipooligosaccharide is required for the generation of infectious elementary bodies in Chlamydia trachomatis', *Proceedings of the National Academy of Sciences*, 108(25), pp. 10284–10289. doi: 10.1073/pnas.1107478108.

Nolan, T., Hands, R. E. and Bustin, S. a (2006) 'Quantification of mRNA using real-time RT-PCR.', *Nature protocols*, 1(3), pp. 1559–82. doi: 10.1038/nprot.2006.236.

O'Connell, C. M. *et al.* (2011) 'Toll-Like Receptor 2 Activation by Chlamydia trachomatis Is Plasmid Dependent, and Plasmid-Responsive Chromosomal Loci Are Coordinately Regulated in Response to Glucose Limitation by C. trachomatis but Not by C. muridarum', *Infection and Immunity*. Edited by S. R. Blanke, 79(3), pp. 1044–1056. doi: 10.1128/IAI.01118-10.

Oliva, G., Sahr, T. and Buchrieser, C. (2015) 'Small RNAs, 5′ UTR elements and RNA-binding proteins in intracellular bacteria: impact on metabolism and virulence', *FEMS Microbiology Reviews*, 39(3), pp. 331–349. doi: 10.1093/femsre/fuv022.

Omsland, A. *et al.* (2014) 'Chlamydial metabolism revisited: interspecies metabolic variability and developmental stage-specific physiologic activities.', *FEMS microbiology reviews*, 38(4), pp. 779–801. doi: 10.1111/1574-6976.12059.

Pages, H. *et al.* (2019) 'Biostrings: String objects representing biological sequences, and matching algorithms.'

Petrovay, F. and Balla, E. (2008) 'Two fatal cases of psittacosis caused by Chlamydophila psittaci.', *Journal of medical microbiology*, 57(Pt 10), pp. 1296–8. doi: 10.1099/jmm.0.2008/001578-0.

Praetzsch, D. (2015) 'Purification of infectious and non-infectious Chlamydia psittaci particles and enrichment of primary transcripts for differential RNA sequencing.', *Friedrich Schiller University (MSc thesis)*.

Rake, G. and Jones, H. P. (1942) 'Studies on lymphogranuloma venereum: I. development of the agent in the yolk sac of the chicken embryo.', *J. Exp. Med.*, 75, pp. 323–338.

Read, T. D. *et al.* (2000) 'Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.', *Nucleic acids research*, 28(6), pp. 1397–406.

Read, T. D. *et al.* (2013) 'Comparative Analysis of Chlamydia psittaci Genomes Reveals the Recent Emergence of a Pathogenic Lineage with a Broad Host Range', *mBio*. Edited by A. Maurelli and P. J. Sansonetti, 4(2). doi: 10.1128/mBio.00604-12.

Ricci, S. *et al.* (1993) 'Transcriptional analysis of the Chlamydia trachomatis plasmid pCT identifies temporally regulated transcripts, anti-sense RNA and sigma 70-selected promoters.', *Molecular & general genetics : MGG*, 237(3), pp. 318–26.

Ritter, J. (1880) 'Beitrag zur Frage des Pneumotyphus (Eine Hausepidemie in Uster [Schweiz] betreffend).', *Deutsches Archiv fuer Klinische Medizin,* (25), pp. 53–96.

Rockey, D. D., Lenart, J. and Stephens, R. S. (2000) 'Genome sequencing and our understanding of chlamydiae.', *Infection and immunity*, 68(10), pp. 5473–9.

Rohde, G. *et al.* (2010) 'Chlamydial zoonoses.', *Deutsches Arzteblatt international*, 107(10), pp. 174–80. doi: 10.3238/arztebl.2010.0174.

Rosenow, C. (2001) 'Prokaryotic RNA preparation methods useful for high density array analysis: comparison of two approaches', *Nucleic Acids Research*, 29(22), p. 112e–112. doi: 10.1093/nar/29.22.e112.

Sachse, K. *et al.* (2009) 'DNA microarray-based genotyping of Chlamydophila psittaci strains from culture and clinical samples.', *Veterinary microbiology*, 135(1–2), pp. 22–30. doi: 10.1016/j.vetmic.2008.09.041.

Saka, H. A. *et al.* (2011) 'Quantitative proteomics reveals metabolic and pathogenic properties of Chlamydia trachomatis developmental forms', *Molecular Microbiology*, 82(5), pp. 1185–1203. doi: 10.1111/j.1365-2958.2011.07877.x.

Saka, H. A. *et al.* (2015) 'Chlamydia trachomatis Infection Leads to Defined Alterations to the Lipid Droplet Proteome in Epithelial Cells.', *PloS one*. Edited by T. Rudel, 10(4), p. e0124630. doi: 10.1371/journal.pone.0124630.

Scheven, M. T. (2014) 'Preparation of high purity infectious and non-infectious particles from Waddlia chondrophila.', *Friedrich Schiller University (MSc thesis)*.

Schmittgen, T. D. and Livak, K. J. (2008) 'Analyzing real-time PCR data by the comparative C(T) method.', *Nature protocols*, 3(6), pp. 1101–8.

Schöfl, G. *et al.* (2011) 'Complete genome sequences of four mammalian isolates of Chlamydophila psittaci.', *Journal of bacteriology*, 193(16), p. 4258. doi: 10.1128/JB.05382-11.

Seth-Smith, H. M. B. *et al.* (2013) 'Whole-genome sequences of Chlamydia trachomatis directly from clinical samples without culture.', *Genome research*, 23(5), pp. 855–66. doi: 10.1101/gr.150037.112.

Seth-Smith, H. M. B. *et al.* (2017) 'The Chlamydia suis Genome Exhibits High Levels of Diversity, Plasticity, and Mobile Antibiotic Resistance: Comparative Genomics of a Recent Livestock Cohort Shows Influence of Treatment Regimes', *Genome Biology and Evolution*, 9(3), pp. 750–760. doi: 10.1093/gbe/evx043.

Sharma, C. M. *et al.* (2010) 'The primary transcriptome of the major human pathogen Helicobacter pylori', *Nature*, 464(7286), pp. 250–255. doi: 10.1038/nature08756.

Sharma, C. M. and Vogel, J. (2014) 'Differential RNA-seq: The approach behind and the biological insight gained', *Current Opinion in Microbiology*. Elsevier Ltd, pp. 97–105. doi: 10.1016/j.mib.2014.06.010.

Sixt, B. S. *et al.* (2013) 'Metabolic Features of Protochlamydia amoebophila Elementary Bodies – A Link between Activity and Infectivity in Chlamydiae', *PLoS Pathogens*. Edited by M. Tan, 9(8), p. e1003553. doi: 10.1371/journal.ppat.1003553.

Sorek, R. and Cossart, P. (2010) 'Prokaryotic transcriptomics: a new view on regulation, physiology and pathogenicity.', *Nature reviews. Genetics*, 11(1), pp. 9–16. doi: 10.1038/nrg2695.

Stephens, R. S., Wagar, E. A. and Edman, U. (1988) 'Developmental regulation of tandem promoters for the major outer membrane protein gene of Chlamydia trachomatis.', *Journal of bacteriology*, 170(2), pp. 744–50.

Swanson, K. A. *et al.* (2009) 'Chlamydia trachomatis polymorphic membrane protein D is an oligomeric autotransporter with a higher-order structure.', *Infection and immunity*, 77(1), pp. 508–16. doi: 10.1128/IAI.01173-08.

Tamura, A. and Higashi, N. (1963) 'PURIFICATION AND CHEMICAL COMPOSITION OF MENINGOPNEUMONITIS VIRUS.', *Virology*, 20(6), pp. 596–604. doi: 14059826.

Tamura, A., Matsumoto, A. and Higashi, N. (1967) 'Purification and chemical composition of reticulate bodies of the meningopneumonitis organisms.', *Journal of bacteriology*, 93(6), pp. 2003–8.

Team, R. D. C. (2013) 'R: A language and environment for statistical computing.', *R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. http://www.R-project.org.*, p. 900051. doi: citeulike-article-id:2400517.

Thomas, V., Casson, N. and Greub, G. (2006) 'Criblamydia sequanensis, a new intracellular Chlamydiales isolated from Seine river water using amoebal co-culture.', *Environmental microbiology*, 8(12), pp. 2125–35. doi: 10.1111/j.1462-2920.2006.01094.x.

Thomason, M. K. *et al.* (2015) 'Global transcriptional start site mapping using differential RNA sequencing reveals novel antisense RNAs in Escherichia coli.', *Journal of bacteriology*, 197(1), pp. 18–28. doi: 10.1128/JB.02096-14.

Thomson, N. R. *et al.* (2005) 'The Chlamydophila abortus genome sequence reveals an array of variable proteins that contribute to interspecies variation.', *Genome research*, 15(5), pp. 629–40. doi: 10.1101/gr.3684805.

Tietzel, I., El-Haibi, C. and Carabeo, R. A. (2009) 'Human Guanylate Binding Proteins Potentiate the Anti-Chlamydia Effects of Interferon-γ', *PLoS ONE*. Edited by D. M. Ojcius, 4(8), p. e6499. doi: 10.1371/journal.pone.0006499.

Vidal, S. *et al.* (2017) 'Neglected zoonotic agents in cattle abortion: tackling the difficult to grow bacteria.', *BMC veterinary research*, 13(1), p. 373. doi: 10.1186/s12917-017-1294-y.

Vijayakumar, S. R. V. *et al.* (2004) 'RpoS-Regulated Genes of Escherichia coli Identified by Random lacZ Fusion Mutagenesis', *Journal of Bacteriology*, 186(24), pp. 8499–8507. doi: 10.1128/JB.186.24.8499-8507.2004.

Voigt, A., Schöfl, G. and Saluz, H. P. (2012) 'The chlamydia psittaci genome: A comparative analysis of intracellular pathogens', *PLoS ONE*, 7(4). doi: 10.1371/journal.pone.0035097.

Volceanov, L. *et al.* (2014) 'Septins Arrange F-Actin-Containing Fibers on the Chlamydia trachomatis Inclusion and Are Required for Normal Release of the Inclusion by Extrusion', *mBio*, 5(5), pp. e01802-14-e01802-14. doi: 10.1128/mBio.01802-14.

Vouga, M., Baud, D. and Greub, G. (2017) 'Simkania negevensis may produce long-lasting infections in human pneumocytes and endometrial cells.', *Pathogens and disease*, 75(1). doi: 10.1093/femspd/ftw115.

Walder, G. *et al.* (2005) 'An Unusual Cause of Sepsis During Pregnancy', *Obstetrics & Gynecology*, 106(5, Part 2), pp. 1215–1217. doi: 10.1097/01.AOG.0000161060.69470.9c.

Wallensten, A., Fredlund, H. and Runehagen, A. (2014) 'Multiple human-to-human transmission from a severe case of psittacosis, Sweden, January-February 2013.', *Euro*

surveillance : bulletin Europeen sur les maladies transmissibles = European communicable disease bulletin, 19(42), pp. 1–6.

Wang, Z., Gerstein, M. and Snyder, M. (2009) 'RNA-Seq: a revolutionary tool for transcriptomics.', *Nature reviews. Genetics*, 10(1), pp. 57–63. doi: 10.1038/nrg2484.

Wehrl, W. *et al.* (2004) 'From the inside out – Processing of the Chlamydial autotransporter PmpD and its role in bacterial adhesion and activation of human host cells', *Molecular Microbiology*, 51(2), pp. 319–334. doi: 10.1046/j.1365-2958.2003.03838.x.

Weinmaier, T. *et al.* (2015) 'Genomic factors related to tissue tropism in Chlamydia pneumoniae infection.', *BMC genomics*, 16, p. 268. doi: 10.1186/s12864-015-1377-8.

Wheelhouse, N. *et al.* (2009) 'Transcriptional analysis of in vitro expression patterns of Chlamydophila abortus polymorphic outer membrane proteins during the chlamydial developmental cycle', *Veterinary Research*, 40(5), p. 47. doi: 10.1051/vetres/2009030.

Yee, M. S. *et al.* (2008) 'Lipoprotein separation in a novel iodixanol density gradient, for composition, density, and phenotype analysis', 49. doi: 10.1194/jlr.D700044-JLR200.

Yu, H. H. Y., Kibler, D. and Tan, M. (2006) 'In silico prediction and functional validation of sigma28-regulated genes in Chlamydia and Escherichia coli.', *Journal of bacteriology*, 188(23), pp. 8206–12. doi: 10.1128/JB.01082-06.

Yu, S.-H., Vogel, J. and Förstner, K. U. (2018) 'ANNOgesic: a Swiss army knife for the RNA-seq based annotation of bacterial/archaeal genomes.', *GigaScience*, 7(9), pp. 1–11. doi: 10.1093/gigascience/giy096.

Zhang, J. *et al.* (2014) 'PEAR: a fast and accurate Illumina Paired-End reAd mergeR.', *Bioinformatics (Oxford, England)*, 30(5), pp. 614–20. doi: 10.1093/bioinformatics/btt593.

8 Curriculum Vitae

Personal information

Name: Thomas Beder
Day of birth: Jul. 17, 1988
Birthplace: Weimar

Career

Since Mar. 01, 2018 **Research associate**
University Hospital Jena
Center for Sepsis Control and Care (CSCC)
Research Topic: “Identification of new targets for insecticides against *Plasmodium falciparum* infected *Anopheles gambiae* using system biological approaches”

Education

2013 – 2019 **Graduation**
Friedrich Schiller University Jena
Faculty of Biological Sciences
Executed at the Leibniz Institute for Natural Product Research and Infection Biology Hans Knöll Institute (HKI), Jena
Department of Cell and Molecular Biology
Supervised by Prof. Dr. Hans Peter Saluz
Title: “Differential RNA-Seq and transcription start site annotation in *Chlamydia*”

2012 – 2013 **Study of Molecular Life Sciences (M.Sc.)**
Friedrich Schiller University Jena
School of Biology and Pharmacy
Mark: 1.2
Master’s thesis
HKI, Jena
Department of Cell and Molecular Biology
Supervised by Prof. Dr. Hans Peter Saluz

Title: “Establishment of basics for the elucidation of primary transcripts of infectious and noninfectious chlamydial particles”
Mark: 1.0

2008 – 2011 **Study of Biology (B.Sc.)**
Friedrich Schiller University Jena
Faculty of Biological Sciences
Mark: 1.9

Bachelor’s thesis
School of Biology and Pharmacy
Plant physiology
Supervised by Prof. Dr. Ralf Oelmüller
Title: “Untersuchung verschiedener ACA (autoinhibited calcium ATPases) Transporter T-DNA Linien auf Homozygotie und Analyse einer Mutante (SALK_010571) in der Interaktion mit *P. indica*.”
Mark: 1.0

2007 – 2008 Mini-job, bike courier at MAILCATS OHG, Weimar

1999 – 2007 Humboldt Grammar School, Weimar
General qualification for university entrance
Mark: 1.8

9 List of publications

Parts of this thesis are included in the following publications:

Beder T, Scheven MT, Praetzsch D, Westermann M, Saluz HP. Purification of infectious and non-infectious chlamydial particles using iodixanol for density gradient preparation. J Microbiol Methods. 2016 Sep;128(3):20–3.

Beder T, Saluz HP. Virulence-related comparative transcriptomics of infectious and non-infectious chlamydial particles. BMC Genomics. 2018 Dec 2;19(1):575.

10 Support

The experimental work for this thesis was performed at the Department of Cell and Molecular Biology (CMB) of the Leibniz Institute for Natural Product Research and Infection Biology – Hans Knöll Institute, Jena. This work was financially supported by the graduate school Jena School for Microbial Communication (JSMC) funded by the German Excellence Initiative [grant number DFG 83/2013].

11 Danksagung

Mein Dank gilt an erster Stelle meinem Doktorvater Prof. Dr. Hans Peter Saluz für die Bereitstellung des spannenden Themas.

Besonders möchte ich mich bei Dr. Gerhard Schöfl für die Einführung in R und NGS-Datenanalyse sowie die ausgezeichnete Betreuung zu Beginn dieser Arbeit bedanken.

Weiterhin gilt mein Dank allen Kollegen und Kolleginnen der Abteilung Zell- und Molekularbiologie, die mich während meiner Arbeit wissenschaftlich unterstützt haben. Im Besonderen möchte ich mich bei Frank Hänel für die Korrekturen und bei Grit Mrotzek, für die hingebungsvolle technische Unterstützung bedanken. Ferner danke ich Mareike Scheven, Dominique Prätzsch und Faisal Khan, die als Studenten tolles Engagement zeigten und für die zahlreichen Stunden mit anregenden Diskussionen.

Weiterer Dank gilt Dr. Martin Westermann vom Elektronenmikroskopischen Zentrum des Universitätsklinikums Jena für die Anfertigung der elektronenmikroskopischen Aufnahmen und Dr. Konrad Sachse vom Friedrich-Loeffler-Institut für die Bereitstellung der Chlamydienstämme.

12 Appendix

Supplementary Table 1. Summary of genome and transcriptome coverages.

Sample			Mapped to genome	Mapped to mRNAs	Avg. read length	Genome coverage	mRNA coverage
C. psittaci	EB_1	-TEX	2,398,313	893,149	132	**270**	**113**
		+TEX	2,838,468	1,458,793	151	**366**	**210**
	EB_2	-TEX	3,769,197	1,570,671	133	**428**	**199**
		+TEX	1,097,803	452,719	137	**128**	**59**
	EB_3	-TEX	4,172,288	1,819,006	136	**484**	**236**
		+TEX	2,783,436	1,117,196	136	**323**	**145**
	EB_4	-TEX	2,989,579	1,394,878	139	**355**	**185**
		+TEX	2,917,537	1,511,777	161	**401**	**232**
	RB_1	-TEX	2,726,118	1,248,206	130	**302**	**155**
		+TEX	2,468,721	982,130	131	**276**	**123**
	RB_2	-TEX	2,799,252	1,622,316	147	**351**	**228**
		+TEX	3,006,588	1,527,173	144	**369**	**210**
	RB_3	-TEX	2,344,889	723,441	137	**274**	**95**
		+TEX	2,299,761	657,519	135	**265**	**85**
	RB_4	-TEX	1,054,294	609,680	135	**121**	**79**
		+TEX	2,054,704	1,280,191	155	**272**	**189**
C. abortus	EB_1	-TEX	3,932,342	607,972	138	**474**	**84**
		+TEX	4,022,348	778,206	137	**482**	**107**
	EB_2	-TEX	4,760,298	431,864	145	**603**	**63**
		+TEX	4,574,202	486,706	135	**540**	**66**
	EB_3	-TEX	2,876,311	1,031,641	126	**317**	**131**
		+TEX	3,509,712	1,565,372	133	**408**	**210**
	RB_1	-TEX	4,116,510	2,929,878	152	**547**	**448**
		+TEX	2,904,883	2,208,272	153	**388**	**340**
	RB_2	-TEX	4,528,435	2,222,027	136	**538**	**304**
		+TEX	1,870,571	1,303,779	152	**248**	**199**
	RB_3	-TEX	3,288,368	2,282,536	150	**431**	**345**
		+TEX	1,754,665	1,399,660	163	**250**	**230**
W. chondrophila	EB_1	-TEX	4,101,177	2,650,252	164	**318**	**223**
		+TEX	3,431,236	1,823,076	156	**253**	**146**
	EB_2	-TEX	4,699,779	2,784,397	156	**348**	**223**
		+TEX	1,170,158	535,060	157	**87**	**43**
	EB_3	-TEX	2,973,025	1,707,053	161	**227**	**141**
		+TEX	2,588,532	1,105,173	154	**189**	**87**
	RB_1	-TEX	753,445	494,351	133	**47**	**34**
		+TEX	1,157,055	671,076	129	**71**	**44**
	RB_2	-TEX	1,208,154	726,397	125	**71**	**47**
		+TEX	1,016,758	549,651	131	**63**	**37**
	RB_3	-TEX	794,860	490,610	139	**52**	**35**
		+TEX	1,832,523	997,408	138	**120**	**71**

Supplementary Table 1. Continued.

E. coli	LB_0.4_1	-TEX	2,022,977	ND	69	**30**	**ND**
		+TEX	2,061,105	ND	78	**35**	**ND**
	LB_0.4_2	-TEX	2,064,583	ND	66	**29**	**ND**
		+TEX	2,063,832	ND	71	**32**	**ND**
	LB_2.0_1	-TEX	3,808,994	ND	64	**53**	**ND**
		+TEX	4,284,902	ND	69	**64**	**ND**
	LB_2.0_2	-TEX	3,964,030	ND	63	**54**	**ND**
		+TEX	5,091,065	ND	70	**77**	**ND**
	M63_0.4_1	-TEX	1,862,651	ND	73	**29**	**ND**
		+TEX	1,834,852	ND	72	**28**	**ND**
	M63_0.4_2	-TEX	1,859,928	ND	61	**24**	**ND**
		+TEX	1,844,238	ND	74	**29**	**ND**
C. penumoniae	EB_1	-TEX	58,642	ND	56	**3**	**ND**
		+TEX	61,770	ND	64	**3**	**ND**
	EB_2	-TEX	29,936	ND	62	**2**	**ND**
		+TEX	46,360	ND	62	**2**	**ND**
	RB_1	-TEX	80,888	ND	64	**4**	**ND**
		+TEX	100,307	ND	67	**5**	**ND**
	RB_2	-TEX	31,506	ND	64	**2**	**ND**
		+TEX	50,191	ND	62	**3**	**ND**

Supplementary Table 2. Summary of gene set enrichement analysis.

		Category	DEG	All	Fisher's exact		G-test		Hypergeometric (enriched)		Hypergeometric (depleted)	
					p-value	padj	p-value	padj	p-value	padj	p-value	padj
C. psittaci	**EB ↑**	AAD	2	24	0.093	0.326	0.044	0.183	0.998	1.000	**0.002**	**0.013**
		CVPGP	9	55	0.162	0.485	0.128	0.383	0.995	1.000	**0.005**	**0.027**
		DNA	5	43	0.072	0.325	0.039	0.183	0.999	1.000	**0.001**	**0.009**
		PMP	0	20	0.012	0.083	0.001	0.010	1.000	1.000	**0.000**	**0.000**
		PROT	87	130	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	1.000	1.000
	RB ↑	AAD	15	24	0.059	0.463	0.058	0.327	**0.003**	**0.040**	0.997	0.997
		CVPGP	28	55	0.066	0.463	0.062	0.327	**0.004**	**0.040**	0.996	0.997
		PROT	20	130	**0.000**	**0.009**	**0.000**	**0.005**	1.000	1.000	**0.000**	**0.000**
		T3A	3	25	0.119	0.624	0.056	0.327	0.997	1.000	**0.003**	**0.036**
	EB +TEX ↑	MT	0	23	0.062	0.562	0.008	0.169	1.000	1.000	**0.000**	**0.000**
	EB +TEX ↓	INC	1	66	0.055	0.382	0.023	0.163	0.995	1.000	**0.005**	**0.026**
		PROT	22	130	**0.001**	**0.025**	**0.001**	**0.023**	**0.000**	**0.001**	1.000	1.000
	RB +TEX ↑	MT	0	23	0.638	1.000	0.097	0.389	1.000	1.000	**0.000**	**0.000**
		PMP	0	20	0.624	1.000	0.122	0.389	1.000	1.000	**0.000**	**0.000**
		RESP	0	21	0.628	1.000	0.113	0.389	1.000	1.000	**0.000**	**0.000**
	RB +TEX ↓	-	-	-	-	-	-	-	-	-	-	-
C. abortus	**EB ↑**	DNA	2	43	0.033	0.158	0.017	0.069	0.998	1.000	**0.002**	**0.009**
		INC	21	55	0.004	0.077	0.004	0.053	**0.000**	**0.004**	1.000	1.000
		T3E	22	65	0.011	0.108	0.010	0.068	**0.001**	**0.010**	0.999	1.000
		T3A	0	23	0.036	0.158	0.005	0.053	1.000	1.000	**0.000**	**0.000**
	RB ↑	T3E	4	65	0.045	0.254	0.021	0.209	0.997	1.000	**0.003**	**0.029**
		T3A	16	23	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	1.000	1.000
	EB +TEX ↑	-	-	-	-	-	-	-	-	-	-	-
	EB +TEX ↓	INC	0	55	0.017	0.333	0.003	0.055	1.000	1.000	**0.000**	**0.000**
		PROT	17	131	0.054	0.442	0.054	0.370	**0.001**	**0.022**	0.999	0.999
		T3E	0	23	0.248	0.825	0.055	0.370	1.000	1.000	**0.000**	**0.000**

Supplementary Table 2. Continued.

		Category	DEG	All	Fisher's exact		G-test		Hypergeometric (enriched)		Hypergeometric (depleted)	
					p-value	padj	p-value	padj	p-value	padj	p-value	padj
C. abortus	**RB +TEX ↑**	AAD	1	22	0.154	0.776	0.074	0.494	0.996	0.997	**0.004**	**0.042**
		T3A	1	23	0.155	0.776	0.063	0.494	0.997	0.997	**0.003**	**0.042**
	RB +TEX ↓	-	-	-	-	-	-	-	-	-	-	-
W. chondrophila	**EB ↑**	CARB	47	81	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	1.000	1.000
		CWC	6	41	0.202	0.485	0.149	0.357	0.987	1.000	**0.013**	**0.039**
		DNA	6	55	0.034	0.137	0.020	0.078	1.000	1.000	**0.000**	**0.002**
		MT	4	37	0.114	0.343	0.056	0.169	0.997	1.000	**0.003**	**0.010**
		MISC	1	22	0.065	0.223	0.023	0.078	0.999	1.000	**0.001**	**0.003**
		PROT	10	156	**0.000**	**0.000**	**0.000**	**0.000**	1.000	1.000	**0.000**	**0.000**
		RESP	23	34	**0.001**	**0.004**	**0.001**	**0.003**	**0.000**	**0.000**	1.000	1.000
		RNA	11	87	0.014	0.066	**0.010**	**0.049**	1.000	1.000	**0.000**	**0.001**
		T3A	19	22	**0.000**	**0.002**	**0.000**	**0.001**	**0.000**	**0.000**	1.000	1.000
	RB ↑	CARB	7	81	**0.000**	**0.000**	**0.000**	**0.000**	1.000	1.000	**0.000**	**0.000**
		PROT	103	156	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	1.000	1.000
		RESP	2	34	**0.003**	**0.019**	**0.001**	**0.008**	1.000	1.000	**0.000**	**0.001**
		RNA	49	87	**0.005**	**0.024**	**0.005**	**0.022**	**0.000**	**0.000**	1.000	1.000
		T3A	0	22	**0.002**	**0.018**	**0.000**	**0.002**	1.000	1.000	**0.000**	**0.000**
	EB +TEX ↑	RESP	1	34	0.112	1.000	0.051	0.950	0.997	1.000	**0.003**	**0.018**
	EB +TEX ↓	-	-	-	-	-	-	-	-	-	-	-
	RB +TEX ↑	T3E	16	205	0.007	0.168	0.005	0.129	0.995	1.000	**0.005**	**0.041**
	RB +TEX ↓	MISC	0	22	0.394	1.000	0.074	0.592	1.000	1.000	**0.000**	**0.000**
		NN	0	25	0.410	1.000	0.057	0.592	1.000	1.000	**0.000**	**0.000**
		RNA	2	87	0.080	1.000	0.038	0.592	0.996	1.000	**0.004**	**0.011**

Differentially expressed genes within each subcategory were tested for over-representation and under-representation using the two-sided Fisher's exact test, G-test and Hypergeometric test and p-values were FDR corrected using the procedure of Benjamini and Hochberg. Categories with fewer than t0w2o genes were omitted in the tables. Shown categories are Amino Acids and Derivatives(**AAD**), Carbohydrates(**CARB**), Cofactors Vitamins Prosthetic Groups Pigments(**CVPGP**), Cell Wall and Capsule(**CWC**), DNA Metabolism(**DNA**), predicted Inclusion Membrane Protein(**INC**), Miscellaneous(**MISC**), Membrane Transport(**MT**), Nucleosides and Nucleotides(**NN**), Polymorphic Membrane Protein(**PMP**), Protein Metabolism(**PROT**), Respiration(**RESP**), RNA Metabolism (**RNA**), T3SS Apparatus(**T3A**) and predicted T3SS Effector(**T3E**).

Supplementary Table 3. Homologous virulence factors (adapted from Collingro *etal.*, 2011)

Function	Gene	*C. psittaci*	*C. abortus*	*W. chondrophila*	*C. trachomatis*	*C. psittaci* EB -TEX vs. RB -TEX		*C.* abotus EB -TEX vs. RB -TEX		*W.* chondrophila EB -TEX vs. RB -TEX	
						L2FC	padj	L2FC	padj	L2FC	padj
Adhesion and host cell entry	ompA	CPS0B_RS00265	CAB_RS00265	-	CT681	-0.81	1.03E-01	0.03	8.99E-01	-	-
	omcB	CPS0B_RS00965	CAB_RS00965	WCW_RS05605	CT443	0.99	2.46E-02	0.22	7.18E-01	-0.67	1.99E-06
	omcA	CPS0B_RS00960	CAB_RS00960	-	CT444	0.51	3.80E-01	-1.06	2.41E-01	-	-
	porb	CPS0B_RS04590	CAB_RS04530	-	CT713	-0.63	1.19E-01	0.76	6.39E-03	-	-
	oprB	CPS0B_RS03655	CAB_RS03590	-	CT372	-0.94	7.84E-03	0.88	1.06E-02	-	-
Effectors secreted by the type III secretion system	pkn5	CPS0B_RS00205	CAB_RS00205	WCW_RS07815	CT673	-0.02	9.52E-01	0.93	1.79E-03	-0.56	5.18E-04
	PknD	CPS0B_RS03395	CAB_RS03325	WCW_RS06335	CT_301	1.39	2.93E-06	0.10	8.48E-01	0.33	7.92E-03
	copB_1	CPS0B_RS04805	CAB_RS04740	-	CT578	0.52	3.80E-01	-0.78	3.43E-01	-	-
	copD_1	CPS0B_RS04810	CAB_RS04745	-	CT579	0.86	1.35E-01	-0.63	4.89E-01	-	-
	copB_2	CPS0B_RS03720	CAB_RS03655	-	CT861	0.70	1.15E-01	0.66	3.92E-03	-	-
	copD_2	CPS0B_RS03725	CAB_RS03660	-	CT860	0.25	5.94E-01	0.10	7.68E-01	-	-
	tarP	CPS0B_RS00890	CAB_RS00890	-	CT456	0.75	1.19E-01	-0.51	5.24E-01	-	-
	cadd	CPS0B_RS05030	CAB_RS04960	-	CT610	-1.63	2.99E-13	-0.09	7.48E-01	-	-
	CT847	CPS0B_RS03800	CAB_RS03740	-	CT847	1.20	1.24E-02	-0.04	9.55E-01	-	-
	mip	CPS0B_RS00420	CAB_RS00425	WCW_RS08630	CT541	-1.06	1.63E-06	-0.41	1.41E-02	0.65	1.97E-04
	incA	CPS0B_RS02790	CAB_RS02780	-	CT119	1.19	1.65E-05	-0.17	4.83E-01	-	-
	incB	CPS0B_RS02485	CAB_RS02475	-	CT232	-1.56	2.27E-09	-0.48	1.63E-02	-	-
	incC	CPS0B_RS02480	CAB_RS02470	-	CT233	-1.71	6.80E-14	-0.33	2.38E-01	-	-
	crpA	CPS0B_RS00970	CAB_RS00970	-	CT442	-0.25	4.23E-01	-2.15	5.40E-22	-	-
	cap1	CPS0B_RS00485	CAB_RS00490	-	CT529	-3.38	1.28E-15	-2.88	2.30E-14	-	-
	DUF582_1	CPS0B_RS00090	CAB_RS00090	-	CT620	0.92	4.73E-02	0.42	3.53E-01	-	-
	SINC	CPS0B_RS00340	CAB_RS00340	-	-	1.05	3.91E-02	-0.24	7.18E-01	-	-
	DUF582_2	CPS0B_RS04600	CAB_RS04540	-	CT711	1.08	4.79E-03	0.18	7.34E-01	-	-
	DUF582_3	CPS0B_RS04595	CAB_RS04535	-	CT712	0.84	6.17E-02	0.40	4.32E-01	-	-
Stress response	nhaD	CPS0B_RS03745	CAB_RS03680	WCW_RS01025	CT857	0.04	8.49E-01	-0.22	2.73E-01	0.56	5.66E-03
	sodA	CPS0B_RS01755	CAB_RS01745	WCW_RS01185	CT294	-1.73	2.32E-14	0.10	7.02E-01	-0.12	5.08E-01
	mgtE	CPS0B_RS02510	CAB_RS02500	WCW_RS04645	CT194	-0.15	3.48E-01	0.08	7.16E-01	-0.20	7.08E-02
	zntA	CPS0B_RS04520	CAB_RS04460	WCW_RS07880	CT727	1.24	1.15E-04	0.13	6.53E-01	-0.24	1.00E-01
	cutA	CPS0B_RS04305	CAB_RS04240	WCW_RS06685	-	3.26	9.72E-06	0.32	4.24E-01	-0.49	1.44E-02
	corC	CPS0B_RS01215	CAB_RS01215	WCW_RS05030	CT423	0.15	5.66E-01	0.25	4.34E-01	0.26	6.84E-02
Drug resistance	phnP	CPS0B_RS01355	CAB_RS01350	WCW_RS05965	CT380	0.33	2.12E-01	0.14	7.16E-01	-0.20	3.92E-01
	marC	CPS0B_RS03770	CAB_RS03710	WCW_RS03655	CT852	1.71	1.20E-09	0.41	3.41E-01	0.83	2.31E-05
Manipulati- on of host cell immune response	cpaF	CPS0B_RS03740	CAB_RS03675	WCW_RS04800	CT858	-1.10	3.62E-10	-1.57	1.13E-14	0.25	2.74E-02
	tsp	CPS0B_RS00975	CAB_RS00975	WCW_RS08025	CT441	1.30	2.33E-03	0.43	2.38E-01	-1.21	4.57E-14
	pgp6-D	CPS0B_RS04830	CAB_RS04765	WCW_RS00095	CT583	0.19	4.77E-01	0.45	5.02E-02	0.20	2.36E-01
Diverse functions	dnaK	CPS0B_RS01240	CAB_RS01240	WCW_RS07900	CT396	-2.80	5.14E-15	-1.08	2.08E-04	0.13	3.57E-01
	hctA	CPS0B_RS04415	CAB_RS04350	WCW_RS00975	CT743	-0.90	1.03E-01	0.75	4.57E-02	-2.00	9.92E-37
	hctB	CPS0B_RS02090	CAB_RS02075		CT046	0.21	6.58E-01	-0.44	4.18E-01	-	-
	groEL	CPS0B_RS03230	CAB_RS03155	WCW_RS06500	CT110	-3.42	4.16E-26	-0.83	5.72E-03	0.66	8.76E-07
	groEL_1	CPS0B_RS04945	CAB_RS04880	-	CT604	0.01	9.62E-01	0.27	4.69E-01	-	-
	groEL_2	CPS0B_RS04355	CAB_RS04290	-	CT755	0.51	6.18E-04	0.51	9.38E-02	-	-
	groES	CPS0B_RS03225	CAB_RS03150	WCW_RS06495	CT_111	-3.63	3.13E-25	-0.62	4.29E-02	0.94	3.10E-08
	htrA	CPS0B_RS03930	CAB_RS03870	WCW_RS02080	CT823	-0.07	6.71E-01	-0.01	9.72E-01	-0.15	3.34E-01
	pmpA	CPS0B_RS01075	CAB_RS01080	-	CT412	0.14	5.90E-01	-0.39	8.59E-02	-	-
	pmpB	CPS0B_RS01070	CAB_RS01075		CT414	-0.38	3.10E-01	0.95	1.77E-03	-	-
	pmpD	CPS0B_RS04045	CAB_RS03975	-	CT812	0.57	9.78E-02	-0.39	1.40E-01	-	-
	pmpE	CPS0B_RS01390	CAB_RS01385	-	CT869	0.07	8.70E-01	1.21	8.94E-06	-	-
	pmpG	CPS0B_RS01410	CAB_RS01405	-	CT871	0.10	7.14E-01	0.76	8.62E-03	-	-
	pmpH	CPS0B_RS01405	CAB_RS01400	-	CT872	0.39	1.86E-01	1.13	1.49E-04	-	-
	pmpI	CPS0B_RS01470	CAB_RS01460	-	CT874	1.13	8.50E-08	-0.26	3.22E-01	-	-
	dsbD2	CPS0B_RS04215	CAB_RS04150	WCW_RS03890	CT780	-1.10	1.67E-04	0.15	5.52E-01	-0.29	1.30E-01
	mviN	CPS0B_RS00070	CAB_RS00070	WCW_RS07555	CT624	1.00	3.52E-04	0.24	4.29E-01	0.34	7.90E-02

L2FC, Log2 fold-change

padj, FDR adjusted p-value

Supplementary Table 4. Categorization of homologous genes in *C. psittaci*, *C. abortus* and *W. chondrophila*

	No. of homologous genes	
Category	**Cpsi-Cab**	**Cpsi-Cab-Wad**
Unknown	158	80
T3SS Effectors	50	14
Inclusion Membrane Proteins	47	4
Cofactors, Vitamins and Prosthetic Groups	20	34
Polymorphic Membrane Proteins	14	0
Amino Acids and Derivatives	10	12
Membrane Transport	9	15
T3SS Apparatus	8	17
RNA Metabolism	6	50
DNA Metabolism	6	37
Fatty Acids, Lipids and Isoprenoids	5	23
Protein Metabolism	4	120
Stress Response	4	5
Genes within Plasticity Zone	4	2
Carbohydrates	3	37
Nucleosides and Nucleotides	2	16
Miscellaneous	2	8
Cell Wall and Capsule	1	30
Respiration	1	20
Virulence, Disease and Defense	0	17
Cell Division and Cell Cycle	0	15
Phosphorus Metabolism	0	4
Sulfur Metabolism	0	2

13 Ehrenwörtliche Erklärung

Die geltende Promotionsordnung der Fakultät für Biowissenschaften der Friedrich-Schiller-Universität Jena ist mir bekannt.

Ich habe die Dissertation selbstständig und ausschließlich unter Verwendung der angegebenen Hilfsmittel, persönlichen Mitteilungen und Quellen angefertigt. Bei der Auswahl und Auswertung des Materials hat mich außer der in der Danksagung genannten Personen niemand unterstützt.

Die Hilfe eines Promotionsberaters wurde nicht in Anspruch genommen und Dritte haben weder unmittelbar noch mittelbar geldwerte Leistungen von mir für Arbeiten erhalten, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen.

Ich versichere, dass ich die Arbeit oder eine in wesentlichen Teilen ähnliche oder eine andere Arbeit bei keiner anderen Hochschule als Dissertation eingereicht habe. Ferner habe ich nicht versucht, die vorliegende Dissertation als Prüfungsarbeit für eine staatliche oder andere wissenschaftliche Prüfung einzureichen.

Jena, den ....

Thomas Beder